A Publication of the American Psychological Association

The Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

Articles

Linguistic Determinism and the Part of Speech 1
Roger W. Brown

A Study of Anxiety in the Sociopathic Personality 6
David T. Lykken

A Study of Affective Responsiveness in a Lie-Detection Situation 11
Jack Block

The Immediate and Delayed Effect of Failure as a Function of Task Complexity and Personalization of Failure 16
Charles B. Truax and Barclay Martin

Uses of Common Objects as Indicators of Cultural Orientations 21
Wayne Dennis

Authoritarianism, Intolerance of Ambiguity, and Rigidity Under Ego- and Task-Involving Conditions 29
Theodore Millon

Cognitive Complexity and Assimilative Projection in Attitude Change 34
Richard M. Lundy and Leonard Berkowitz

Make-A-Picture Story (MAPS) Test Findings for Rubella Deaf Children ... 38
D. Marjorie Bindon

Continued on inside front cover

Volume 55 JULY 1957 Number 1

CONTENTS—*Continued*

Shift in Judgment of Weights as a Function of Anchoring Stimuli and Instructions in Early Schizophrenics and Normals 43
Kurt Salzinger

An Analysis of Interperson Correlations Among Thirty Psychotics 50
Robert J. Williams and Vincent S. Machi

Frequency and Idiosyncracy of Associative Responses 58
George Mandler and Elizabeth W. Parnas

The Relation of Ego Control to Overt Aggression and Dependency 66
Norman Livson and Paul H. Mussen

Attitude Change Through Reward of Verbal Behavior 72
William Abbott Scott

Identification and Observational Learning from Films 76
Eleanor E. Maccoby and William Cody Wilson

Body-Image Boundaries and Adjustment to Poliomyelitis 88
Kenneth E. Ware, Seymour Fisher, and Sidney Cleveland

The Manifest Anxiety Scale as a Measure of Drive 94
Robert E. Silverman

The Psychological Meaning of Acquiescence Set for Authoritarianism 98
N. L. Gage, George S. Leavitt, and George C. Stone

Cultural Influence on the Perception of Movement: The Trapezoidal Illusion Among Zulus 104
Gordon W. Allport and Thomas F. Pettigrew

A Motivational Factor in Conformity 114
George Moeller and Mortimer H. Applezweig

Group Learning and Communication as a Function of Task and Structure "Demands" 121
John T. Lanzetta and Thornton B. Roby

CRITIQUE AND NOTES

A Further Study of the Effect of Stress on Palmar Prints 132
Robert H. Davis

Situational Differences in Leader Behavior 132
A. Paul Hare

Religious Practice and Cardiovascular Reactions During Stress 135
Stanley H. King and Daniel M. Funkenstein

Certainty of Judgment and Resistance to Social Influence 137
Harold H. Kelley and Thomas W. Lamb

Two Kinds of Assumed Similarity Between Opposites 140
Ivan D. Steiner and Colin G. McDiarmid

Castration, Circumcision, and Anti-Semitism 143
Irving E. Alexander and Sheldon Blackman

Consistency in Response to Group Pressures 146
Stanley Rosner

On the Relations Between Discrimination Reaction Time, Anxiety, and Intelligence 147
G. Robert Grice

Security of Judges as a Factor in Impressions of Warmth in Others 147
Joseph Bossom and A. H. Maslow

Volume 55 JULY, 1957 Number 1

The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *New York University*

Associate Editor: E. J. SHOBEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
The RAND Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

LEON FESTINGER
Stanford University

DONALD W. FISKE
University of Chicago

ROBERT R. HOLT
New York University

HERBERT H. HYMAN
Columbia University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
Center for Advanced Study in the Behavioral Sciences

LEONARD S. KOGAN
Community Service Society of New York

GARDNER LINDZEY
Syracuse University

DAVID C. MCCLELLAND
Harvard University

QUINN MCNEMAR
Stanford University

LEO POSTMAN
University of California

ELIOT H. RODNICK
Duke University

NEVITT SANFORD
University of California

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

RENATO TAGIURI
Harvard University

Managing Editor: ARTHUR C. HOFFMAN

Assistant Managing Editor: HELEN ORR

Editorial Staff: BARBARA CUMMINGS, FRANCES H. CLARK, SADIE J. DOYLE

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 784 pages. The subscription price per year is $16.00; foreign, $16.50. Single issues are $3.00.

Original manuscripts should be sent to the Editor:

DR. M. BREWSTER SMITH
Graduate Department of Psychology
New York University
Washington Square
New York 3, N. Y.

Fifty free offprints are supplied to contributors. Authors of prior publication articles receive no gratis offprints.

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Printed and mailed by:

WAVERLY PRESS, INC.
Baltimore 2, Maryland

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Mt. Royal and Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Entered as second-class matter February 15, 1954 at the Post Office at Baltimore, Maryland under the Act of February 28, 1925. Acceptance for mailing at special rate of postage provided for in Section 34.40 (e), P. L. and R. Additional entry at Washington, D.C.

Copyright 1957 by the American Psychological Association, Inc.

LINGUISTIC DETERMINISM AND THE PART OF SPEECH[1]

ROGER W. BROWN
Harvard University

IN RECENT years the anthropologists Whorf (12), Lee (9), and Hoijer (6) have put forward the view that language is a determinant of perception and thought. The nature of the determining influence exerted by the vocabulary of a language is quite clear (1), but it is less easy to see how the grammatical features of a language can affect cognition. Yet it is just the grammatical differences between languages that are most striking and it is their determining force that the anthropologist has stressed. This paper undertakes to show how one kind of grammatical practice, the allocation of words to one or another part of speech, does affect cognition.

The words of a language can be collected into classes of formal equivalents which are called the parts of speech. Fries (5) has shown that English nouns, for example, are words acceptable in sentence frames of the type "(The) ——— is (are) good." Native speakers of English will find it possible to insert "concert" or "cow" or "truth" in that frame but will find "very" or "of" grammatically impossible. Fries has used other sentence frames to separate out verbs, adjectives, and adverbs. In French, nouns can be further subdivided into formal gender classes. In Navaho there are more than 20 formal classes for words naming different kinds of objects. All of these word classes are defined by linguistic science in terms of the combinational possibilities of forms in a language without reference to the meanings of forms.

So long as these classes are defined in purely formal terms they do not suggest important cognitive differences. That suggestion comes in when we add the semantic correlates of the classes. The native speaker of English is likely to think of the parts of speech in semantic terms. Nouns name substances; verbs name processes; and adjectives name qualities. The genders of certain European languages are usually called masculine, feminine, and neuter, and these are semantic characterizations. The object classes of Navaho are usually described as words naming round objects, words naming long, thin objects, words naming granular substances, etc. The linguistic determinists in anthropology believe that the semantic character of the form classes fixes the fundamental conception of reality in a language community and that differences on this level correspond to different Weltanschauungen.

At the same time the science of descriptive linguistics refuses to define its word classes in semantic terms. Fries (5) has shown that for the English parts of speech such definitions are always either unclear or overextended. We all know the English teacher's characterization of the noun as the name of a person, place, or thing. The terms "person" and "place" are reasonably clear, but do not apply to such nouns as "truth," "odor," and "thought." The meaning of "thing" is so unclear that we cannot tell whether or not odor and thought are things. No one has been able to provide clear semantic definitions that will serve to distinguish every English noun from every verb, adjective, and adverb. It is well known that the "masculine" and "feminine" genders in the European languages include names for objects having no sex. In Navaho, too, the object classes do not show perfect semantic consistency. In short, the semantic definitions of the form classes ignore many exceptions and are unsuitable for the purposes of linguistic science.

When the linguistic scientist sets up his descriptive categories he quite naturally looks for attributes of exceptionless validity, and there are not such semantic attributes for the English parts of speech. However, the layman may operate, in this area as in so many others, with conceptions that take account of probabilistic as well as certain associations. It may be that nouns *tend to have* a different semantic from verbs, and that the native speaker detects this tendency while he is in the process of learning the language. To answer these questions examination was made of the nouns and verbs of young children learning English to see whether there was a semantic distinction be-

[1] The gist of this paper was presented in a conference on "Linguistic meaning" sponsored by the Social Science Research Council and held at Yale University, May 17–18, 1956. The Laboratory of Social Relations, Harvard University, financed the research.

tween the two parts of speech. The distinction proved to be much clearer than it is in the vocabulary of English-speaking adults. The second step was an experiment to find out whether the children were aware of the semantic distinction between nouns and verbs and whether they made any use of the distinction.

THE NOUN AND VERB IN CHILDREN'S SPEECH

Harvard pre-school sessions were visited for about a month. There were eight children in each class; two of the classes were limited to children between four and five years while a third class accepted those between three and four. As an observer, the author sat on the side-lines and let the pre-school life swirl about him, recording verbatim all the conversation he could hear. From these records, he made vocabulary lists classified into the parts of speech. It was his impression, on examining this vocabulary, that the nouns and verbs of children were more nearly consistent with the classroom semantic definitions than are the nouns and verbs of adults. Nouns commonly heard were "truck," "blocks," and "teacher." There were no uses of "thought" or "virtue" or "attitude." These observations suggested that as the form classes grow larger they decline in semantic consistency. Perhaps children develop firm, and temporarily reliable, notions about the semantics of nouns and verbs. These notions may stay with them as adults even though they retain only a probabilistic truth.

To compare the character of adult and child vocabularies, the first thousand most frequent words from the Thorndike-Lorge (11) list of adult usage were examined, and also the first thousand most frequent words from the Rinsland (10) list of the vocabulary of children in the first grade. The Rinsland list is based on 4,630 pages of conversation, plus more than a thousand letters and stories. The Rinsland list is much the same as lists compiled independently by the Child Study Committee of the Kindergarten Union (8) and by Horn (7).

The first set of contrasts deals with two reduced lists; nouns found among the first thousand for adults but not for children, compared with nouns among the first thousand for children but not for adults. The set of nouns having clearest "thing" character would seem to be those that are called "concrete" and it is a commonplace to describe the language of children as more concrete than that of adults. One sense of the pair "concrete-abstract" is the same as "subordinate-superordinate." The more abstract term, the superordinate, includes in its denotation the denotation of the concrete or subordinate term, but extends beyond it. Superordinate-subordinate relations between the two lists were all in one direction. The adult list included "action," "article," "body," "experience," and at least seven others which were superordinate to many words on the children's list. There were no nouns on the children's list superordinate to those on the adult list.

The concrete noun with the smaller denotation is likely to be more picturable than its superordinate, and picturability is another common sense of "concrete." Of course the concrete noun, like the abstract, names a category rather than a particular instance. However, some categories have a more or less characteristic visual contour and size while others do not. Visual contour is a defining attribute for "table," but not for "thing" or "experience." Of the adult nouns, 16% named categories having a characteristic visual contour, while 67% of the children's nouns were of this kind. Nouns like "apple," "barn," and "airplane" name categories for which size is a defining attribute, while nouns like "affair," "amount," and "action" do not. On the adult list, 39% of the nouns were of the former kind, while 83% of the children's nouns had size implications. It appears that children's nouns are more likely to name concrete things (in the sense of naming narrow categories with characteristic visual contour and size) than are the nouns of adults.

Two lists of verbs were compared: those among the first 1,000 for adults but not for children, and those among the first 1,000 for children but not for adults. The question here was the percentage of verbs naming animal (including human) movement. Of the adult verbs, 33% were of this kind, while 67% of the children's verbs named actions. The common notion that verbs name actions seems to be truer for the vocabulary of children than for the vocabulary of adults.

These studies of word lists confirm the impression that the nouns and verbs used by children have more consistent semantic implications than those used by adults. It remains

a question whether children are, in any sense, aware of these implications. There are many ways in which such awareness could be useful to one learning the language. Adults often try to convey the sense of a word by speaking it in the presence of the object or event named. All such single namings are ambiguous. The adult who says "water" while looking at a glass of water may cause a child to attend to the glass itself as a container, to the glass as a transparent material, to the liquid character of its contents, to the height of the liquid, to the state of containment, and so on. Selection of the nonlinguistic attributes that govern proper denotative use of the word "water" cannot be guaranteed by a single naming. Repeated pointings can, of course, establish the invariant circumstances governing use of the word. If there were nothing to suggest to the child the probably relevant features of the nonlinguistic world, discovery of linguistic meanings would be a very laborious affair. However, a new word is ordinarily introduced in a way that makes its part-of-speech membership clear: "Look at the *dog*" or "See him *running*." If a part of speech has reliable semantic implications it could call attention to the kind of attribute likely to belong to the meaning of the word. A child who had absorbed the semantics of the noun and verb would know, the first time he heard the word "dog," that it was likely to refer to an object having characteristic size and shape, whereas "running" would be likely to name some animal motion. The part-of-speech membership of the new word could operate as a filter selecting for attention probably relevant features of the nonlinguistic world. It seemed that one could learn whether children experience any such filtering of attributes by introducing to them newly invented words assigned to one or another part of speech, and then inquiring about the meanings the words appeared to have.

In the children's speech that had been recorded, nouns and verbs were given proper grammatical treatment. In addition, the children made correct use of a subclass of nouns—the mass nouns. These are words like "dirt," "snow," "milk," and "rice" which are given different grammatical treatment from such particular nouns as "barn," "house," and "dog." For example, when "some" is used with "barn" the noun is in the plural—"some barns," whereas a mass noun would be in the singular—"some rice." The semantic difference between these two classes of noun is suggested by the designations "mass" and "particular." Mass nouns usually name extended substances having no characteristic size or shape, while particular nouns name objects having size and shape. Many nouns can function in either a mass or particular way with attendant shifts in the speaker's view of the referent. "Some cake" is a chunk of a mass while "some cakes" are either cupcakes or layer cakes arranged in a row. Many words in the vocabulary of psychology have this double potentiality. Although the personologist deplores such usage, the layman speaks of someone having "a lot of personality" or "very little temperament." The professional insists that personality is not an undifferentiated substance of which one can have more or less. Personalities are like cupcakes—all of a size and one to a customer—with only their frostings to make them unique.

In the speech of the pre-school children "milk," and "orange juice," and "dirt" were the most common mass nouns. These were always given correct grammatical treatment. No one said "a milk" or "some dirts." It was decided to work with three functional classes: the particular noun, the mass noun, and the verb.

METHOD

The experiment involved three sets of four pictures each.[2] One of these sets will be described in detail. The first picture in the set shows a pair of hands performing a kneading sort of motion, with a mass of red confetti-like material which is piled into and overflowing a blue-and-white striped container that is round and low in shape. The important features of the picture are the kneading action, the red mass, and the blue-and-white round container. The motion would ordinarily be named with a verb (like "kneading"), the mass with a mass noun (like "confetti"), and the container with a particular noun (like "container"). It was assumed that children would have no readily available names for any of these conceptions. Each of the remaining three pictures of this set exactly reproduced one of the three salient features of the first picture, either the motion, the mass, or the container. In order to represent the motion a second time it was necessary to show also a mass and a container. However, the mass was here painted yellow so as not to duplicate the original, and the container was of a different size, shape, and color from the original. The other two sets of pictures involved different content, but always an action, a mass

[2] The author is grateful to Dr. Susan Ervin for painting the pictures and for suggesting colors and forms that would please children.

substance, and a particular object. In one case, the first picture showed hands cutting a mass of cloth with a strange tool. In the third set, hands were shown emptying an odd container of a slushy sort of material.

In overview, the following use was to be made of the three sets of pictures. Children were to be shown the first picture in conjunction with a new word identifiable either as a verb, a mass noun, or a particular noun. Then they would be shown the remaining three pictures of the set and asked to point out the one that pictured again what had been named in the first picture. It was anticipated that when the new word was a verb they would point to the picture of motion, when it was a particular noun they would point to the container, and when it was a mass noun they would point to the extended substance.

Three word stems were used: "niss," "sib," and "latt." If the stem was to function as a verb X would begin by asking: "Do you know what it means to sib?" (Children do not always answer "no" as they ought.) "In this picture" (first picture of a set) "you can see sibbing. Now show me another picture of sibbing" (presenting the other three pictures of the set). If the stem was to function as a particular noun, X began: "Do you know what a sib is?" and proceeded in consistent fashion. If the word was to function as a mass noun, X began: "Have you ever seen any sib?" and went on accordingly.

Each child saw all three sets of pictures and heard each of the word stems; one of them as a particular noun, one as a mass noun, and one as a verb. The combinations of word stem, part-of-speech membership, picture set, and order of presentation were all randomly varied. There were 16 children in all, half of them between three and four years, and half between four and five. They were all acquainted with X by the time the experiment was performed. The procedure was very like the familiar business of looking at a picture book and naming the things seen and was accepted by the children as a kind of game. The game was always played with one child at a time.

RESULTS AND DISCUSSION

When a new word was introduced as a verb, 10 of the 16 children picked out the picture of movement. When the word was a particular noun, 11 of 16 selected the picture of an object; and when the word was a mass noun, 12 of 16 selected the extended substance. Of the 15 responses that were not correct, four were simply failures to answer because of some distraction from the task. The results are summarized in Table 1. A simple test was made to determine the significance of the differences in the pictures selected by children when the new word was a verb, when the word was a particular noun, and when it was a mass noun. For example, the selections made when the word was a verb were dichotomized into pictures of actions and all others. These frequencies for verbs were compared with like frequencies for the choices made when the words were either particular or mass nouns. In other words, the test was to determine whether action pictures were more likely to be selected as referents for new words introduced as verbs than for new words introduced as nouns. Comparable tests were made to see whether particular nouns were associated with pictures of objects and mass nouns with substances. All three of the resultant 2×2 tables yielded differences beyond the .005 level of significance when the Fisher-Yates test was applied.

TABLE 1

PICTURE SELECTIONS FOR WORDS BELONGING TO VARIOUS PARTS OF SPEECH

Category Depicted	Verbs	Particular Nouns	Mass Nouns
Actions	10	1	0
Objects	4	11	3
Substances	1	2	12
No response	1	2	1

It is well known that children will sometimes do what an adult wishes in a task of this kind though they do not understand the task as the adult does. Consequently, the qualitative results may be more persuasive than the quantitative. In the first trial with the first child, for instance, X showed the picture of cloth being cut by an odd tool and said that there was a "sib" in the picture. Then went on with: "Can you show me another sib?" and while X still fumbled with the other three pictures, his subject swung around and pointed to the steam valve on the end of the radiator saying, "There's a sib." The pictured tool looked very like the steam valve. In another case, X showed the picture of confetti-kneading and said, "There is some latt in this picture," whereupon his subject said: "The latt is spilling." And it was.

Recent experiments with phonetic symbolism (2) and metaphor (3) indicate that semantic rules are not always arbitrary. A word can suggest its meaning because the sound is an echo of the sense or because the word had a prior meaning which is related to the new meaning. The present study suggests that most words have an additional kind of "appropriateness" stemming from their grammatical character. While the part-of-speech membership of a word does not give away the particu-

lar meaning, it does suggest the general type of that meaning, whether action, object, substance, or whatever. In learning a language, therefore, it must be useful to discover the semantic correlates of the various parts of speech; for this discovery enables the learner to use the part-of-speech membership of a new word as a first clue to its meaning. The present experiment with very young children who are learning English indicates that in this language, at least, the semantic implications of the verb, mass noun, and particular noun are discovered by native speakers. It now seems quite probable that speakers of other languages will also know about the semantics of their grammatical categories. Since these are strikingly different in unrelated languages, the speakers in question may have quite different cognitive categories. It remains to be determined how seriously and how generally thought is affected by these semantic distinctions.

SUMMARY

Descriptive linguistics defines the parts of speech in strictly formal or syntactical terms. Nevertheless, the parts of speech usually have distinct semantic characteristics. These characteristics do not hold for all members of the various parts of speech, however, and so cannot serve to define the parts of speech for the purposes of linguistic science. Human beings are generally adept at picking up imperfect probabilistic implications, and so it may be the case that native speakers detect the semantic nature of the parts of speech of their language. It was shown that the nouns used by young English-speaking children were more reliably the names of things and their verbs more reliably the names of actions than is the case for the nouns and verbs used by English-speaking adults. It was shown experimentally that young English-speaking children take the part-of-speech membership of a new word as a clue to the meaning of the word. In this way, they make use of the semantic distinctiveness of the parts of speech. It seems likely that speakers of languages other than English will also have detected the semantic characters of their parts of speech. There is a sense, then, in which this grammatical feature of a language affects the cognition of those who speak the languages. Differences between languages in their parts of speech may be diagnostic of differences in the cognitive psychologies of those who use the languages.

REFERENCES

1. BROWN, R. W. Language and categories. Appendix to J. S. Bruner, J. J. Goodnow, & G. A. Austin, *A study of thinking*. New York: Wiley, 1956.
2. BROWN, R. W., BLACK, A. H., & HOROWITZ, A. E. Phonetic symbolism in natural languages. *J. abnorm. soc. Psychol.*, 1955, **50**, 388–393.
3. BROWN, R. W., LEITER, R. A., & HILDUM, D. C. Metaphors from musical criticism. *J. abnorm. soc. Psychol.*, **54**, 347–352.
4. BUCKINGHAM, B. R., & DOLCH, E. W. *A combined word list*. Boston: Ginn, 1936.
5. FRIES, C. C. *The structure of English*. New York: Harcourt Brace, 1952.
6. HOIJER, H. (Ed.) *Language in culture*. Chicago: Univer. of Chicago Press, 1954.
7. HORN, E. The commonest words in the spoken vocabulary of children up to and including six years of age. In National Society for the Study of Education, *Twenty-fourth yearbook*. Bloomington, Illinois: Public School Publishing Company, 1925.
8. INTERNATIONAL KINDERGARTEN UNION, CHILD STUDY COMMITTEE. *A study of the vocabulary of children before entering the first grade*. Washington, D. C.: The Int. Kindergarten Union, 1928.
9. LEE, DOROTHY D. Conceptual implications of an Indian language. *Phil. of Sci.*, 1938, **5**, 89–102.
10. RINSLAND, H. D. *A basic vocabulary of elementary school children*. New York: Macmillan, 1945.
11. THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Bur. of Publ., Teachers College, Columbia Univer., 1944.
12. WHORF, B. L. *Language, thought, and reality: selected writings*. Cambridge: Technology Press, 1956.

Received August 17, 1956.

A STUDY OF ANXIETY IN THE SOCIOPATHIC PERSONALITY[1]

DAVID T. LYKKEN

University of Minnesota

THE concept of the psychopathic personality includes so heterogeneous a group of behavior disorders as to be at least two steps removed from the level of useful psychiatric diagnosis. Sociopathic personality is a more recent designation (1) which refers to a subgroup of these disorders in which the pathognomic characteristics are impulsiveness, antisocial tendencies, immorality, and a seemingly self-destructive failure to modify this pattern of behavior in spite of repeated painful consequences. This category may be regarded as a genus composed of phenotypically similar, but etiologically distinct, subtypes such as the dissocial and the neurotic sociopaths.

A third species has been described (3, 12, 13, 14, 17), which may be called *primary sociopathy*, in which neither neurotic motivations, hereditary taint, nor dissocial nurture seem to be determining factors. Cleckley (3) has reported the chief clinical characteristic of this group as a lack of the normal affective accompaniments of experience. If this observation is correct, it would point the way toward accurate diagnostic isolation of primary sociopathy as well as guiding research into the question of its etiology. Classification according to the presence or absence of defective emotional reactivity, therefore, satisfies one criterion of useful diagnosis in that it shows promise of relationship to the as yet unknown origins of the disorders to be distinguished.

The other requirement for useful diagnosis is that the criteria of classification must be objective. Clinical assessment of the "normality of the affective accompaniments of experience" is subjective and unreliable. In consequence, Cleckley's work has had as yet little real impact on psychiatric practice. By expressing this putative defect of the primary sociopath in terms of the anxiety construct of experimental psychology (18, 19, 20, 21, 22), it becomes susceptible to quantification and empirical test.

An experimental hypothesis may now be formulated. Among persons conventionally diagnosed as psychopathic personality, those who closely resemble the syndrome described by Cleckley are (*a*) clearly defective as compared to normals in their ability to develop (i.e. *condition*) anxiety, in the sense of an anticipatory emotional response to warning signals previously associated with nociceptive stimulation. Persons with such a defect would also be expected to show (*b*) abnormally little *manifest anxiety* in life situations normally conducive to this response, and to be (*c*) relatively incapable of *avoidance learning* under circumstances where such learning can only be effected through the mediation of the anxiety response.

METHOD

The Sample

The extreme heterogeneity, even on the crudest descriptive level, of persons diagnosed as psychopathic personalities in various clinical or institutional settings complicated the selection of an appropriate experimental sample. The institution psychologists[2] were given a list of 14 criteria drawn from Cleckley (3, pp. 355–392) and were asked to compare against these criteria those inmates diagnosed as psychopathic personality. Inmates who, in their opinion, best fitted the Cleckley prototype were listed as candidates for experimental Group I, the primary sociopathic group. Inmates who they felt did *not* meet the criteria in important respects were listed as candidates for experimental Group II, designated as the neurotic sociopathic group. In this selection process, the psychologists were asked to reaffirm the original diagnosis, discarding from consideration for either group those inmates who, in their present opinion, would not be diagnosed as psychopathic personality at all.

A control Group III of 15 "normals," roughly comparable in age, intelligence, and socioeconomic background, was selected from the University General College and a local high school.

Group I, composed of 12 males and 7 females, had a

[1] Drawn from a thesis submitted to the University of Minnesota in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is indebted to his adviser, Professor Ephraim Rosen, and to others whose assistance aided in the completion of this research.

[2] The writer is indebted to the administrators and to the psychologists of the Minnesota State Reformatory, St. Cloud, Minnesota; the Minnesota State Reformatory for Women, Shakopee, Minnesota; the State Home for Girls, Sauk Centre, Minnesota; and the St. Peter State Hospital, St. Peter, Minnesota.

mean age of 21.6 years (SD = 4.3), and a mean IQ of 109.2 (SD = 10.7). Group II included 13 males and 7 females, had a mean age of 24.5 years (SD = 5.4), and a mean IQ of 104.5 (SD = 8.8). For the 10 male and 5 female normals, the mean age was 19.07 (SD = 3.2), and the mean IQ 100.4 (SD = 10.2). None of these group differences were significant.

The Measures and Testing Procedure

It was necessary to do the testing at the several institutions under varying conditions. In all cases, however, the apparatus was arranged on a large table, the experimenter on one side and the subject (S) seated comfortably opposite. The S was told that he was assisting in a psychological experiment having no bearing on his personal record and that his performance would be treated with strict anonymity. An attempt was made throughout to keep the testing on an informal basis.

As an indicant of manifest anxiety as referred to in hypothesis *b*, an "Anxiety Scale" was constructed expressly for this study to supplement the Taylor scale and Anxiety Index which appear to be more strongly loaded on a factor of neurotic self-description. In this new scale, each of the thirty-three items involves two statements of activities or occurrences, matched for general unpleasantness or undesirability according to a modified Thurstone scaling procedure utilizing 15 college student judges. One activity of each pair is unpleasant, presumably because of its frightening or embarrassing character (e.g., "making a parachute jump" or "knocking over a glass in a restaurant"). The paired activity is intended to be onerous but not frightening (e.g., "digging a big rubbish pit" or "cleaning up a spilled bottle of syrup"). The S is required to choose that member of each pair which he would prefer as a lesser of evils. The degree to which the "frightening" alternatives are rejected is interpreted as an index of the extent to which anxiety determines behavior choices within the range of life situations sampled by this test.

The booklet form of the MMPI was used and the answer sheets scored and K-corrected in the usual way (10). The Anxiety Index, or AI, was calculated according to the formula given by Welsh (23). The Heineman form (11) of the Taylor scale was given and scored by subtracting the number of "anxiety" items rejected as "least applies to me" from the number endorsed as "most applies to me."

An avoidance learning test was given to determine whether there were group differences in capacity to learn on the basis of anxiety reduction (hypothesis *c*). It involved an elaborate, electrically operated mental maze which the S was given 20 trials to learn (the "manifest task"). At each of the 20 choice points in this maze, choice of one of the 4 possible alternatives (always an error alternative) gave an electric shock. It was intended that social and ego rewards should reinforce performance in the manifest task. Performance on the "latent task," which was to avoid the shocked alternatives—to err instead on the unshocked alternatives—was presumably reinforced only through anxiety reduction.

The measure of anxiety conditionability (hypothesis *a*) employed the GSR as the dependent variable. A shocking electrode was attached to S's nondominant hand, the GSR electrodes being already in place on the dominant hand. The S was told that after the blindfold had been replaced, he was to sit as quietly as possible for the next 30 to 40 minutes, during which time he would periodically hear a buzzer (which was then demonstrated) and occasionally receive a brief electric shock. When the S was seated comfortably and relaxed insofar as possible, the recording apparatus was started and the conditioning series (CS) begun.

Two buzzers were used which were distinguishably different in timbre rather than in pitch, the difference being one not easily labeled (to minimize verbal mediation of a discrimination between them). Buzzer No. 1 was used as the CS and was the only one reinforced; buzzer No. 2 was used to test for generalization effects. In all cases, stimuli of the conditioning series were presented as soon as GSR activity from preceding stimuli had subsided, the intertrial interval being therefore not constant within or between Ss, but averaging between 20 and 60 seconds. (This method of stimulus timing automatically eliminates temporal conditioning.) When turned on, the buzzers sounded for a period of 5 seconds, controlled by an automatic timer.

The reinforcing stimulus or UnCS was an electric shock from a 700-volt AC supply through two 68,000-ohm series resistors, presented automatically for about 100 milliseconds just before the termination of the CS (buzzer No. 1). The shock was applied between an electrode on the palm of one hand and the GSR *ground* electrode on the palmar tip of the middle finger of the opposite hand. The shock sensation was felt mainly on the richly innervated finger tip and was a decidedly unpleasant stimulus, producing in most cases a pronounced startle reaction and in all cases a strong GSR.

The sequence of trials or stimulus presentations was as follows:

1. To permit the adaptation of unconditioned GSR to the buzzers themselves, stimuli were first presented without shock reinforcement for a total of 10 trials in the order 2, 1, 2, 1, S, 2, 2, 1, 1, 1, 1. A single preliminary shock was given in the series at the point S, separated by at least 30 seconds from the buzzers occurring before and after it.

2. Seven consecutive shock-reinforced presentations of the CS were given as the conditioning series, followed by four more reinforcements interspersed with four unreinforced trials with buzzer No. 2 in the order 1, 1, 1, 1, 1, 1, 2, 1, 2, 2, 1, 2, 1, 1.

3. A total of 24 extinction trials was then given, the two buzzer stimuli being presented in the order 1, 2, 1, 1, 2, 1, 2, 1, 1, 2, 1, 1, 1, 1, 2, 2, 1, 2, 1, 1, 2, 1, 1, 1. Considering only the CS, buzzer No. 1, the series therefore consisted of 6 prereinforcement trials, 11 reinforced conditioning trials, and 16 extinction trials.

Skin resistance was measured by a modification of a circuit suggested by Flanders (6) which passed an electronically regulated constant DC current of 40 microamperes through S. The electrodes were curved discs of Monel metal, 15 mm. in diameter, applied to the palmar surface of the distal phalange of the first, second, and third fingers of the same hand. The skin surface was first scrubbed with alcohol and then coated with Sanborn electrode paste. The exosomatic current was applied between the first and third fingers, which were

also connected to the push-pull input grids of a Sanborn Model 126 DC amplifier, driving a Sanborn Model 127 recording milliameter. The electrode on the second finger was connected to amplifier and external ground. The instrument was calibrated before each use and provided a linear record of resistance and resistance change, accurate to less than ±50 ohms.

All GSRs were recorded in terms of resistance change. A variety of transformations was then applied and tested against the usual criteria of normality of distribution, correlation with basal resistance, and homogeneity of variance across people with respect to several test stimuli (2, 8, 9, 16). The result of this analysis was that each resistance change was expressed as the logarithm of the ratio of that change to the mean resistance change produced by the first six electric shocks. This unit expresses the galvanic CR as a proportion of the individual's UnCR and, for a conditioning study, seems quite appropriate for individual comparisons.

Three GSR indices were derived from the protocols of the conditioning series: (*a*) GSR Reactivity, which is the mean GSR to the CS during the fourth through seventh conditioning trials; (*b*) GSR Conditioning, which is equal to (*a*) minus the mean GSR to the last three preconditioning trials and the last three extinction trials (this index measures essentially the slope of the conditioning curve or the increment actually produced by the reinforced trials); (*c*) GSR Generalization, the ratio of the mean GSR to buzzer No. 2 during early extinction trials 18, 20, 21, 23 to the mean GSR to buzzer No. 1 during trials 17, 19, 22, 24.

The testing sequence was as follows: (*a*) Anxiety scale; (*b*) GSR Conditioning series; (*c*) Avoidance Learning test; (*d*) MMPI (given during the week following the foregoing individual testing); (*e*) Taylor Manifest Anxiety Scale, forced-choice form given later with the MMPI.

RESULTS AND DISCUSSION

Scores on all measures were converted for easier comparison to a standard score form with each distribution having a grand mean of 500 and a standard deviation of 100. Group means on all measures, together with significance test results, are given in Table 1.

TABLE 1
GROUP MEANS ON ALL MEASURES: SIGNIFICANCE TESTS*

Measure	Group			*d*-Test
	I	II	III	prob.†
Taylor Scale	471	556	462	.01
Anxiety Index	472	557	464	.01
Anxiety Scale	470	511	529	.05
MMPI *Pd* Scale	532	547	395	.05
Avoidance Learning	461	501	558	.01
GSR Reactivity	498	494	534	.05
GSR Conditioning	478	483	551	.05
Generalization	473	542	490	—

* All measures converted to a scale having an over-all mean of 500 and *SD* of 100.

† Probabilities given are for significance of largest difference (e.g., III − I for GSR Conditioning). Significance test was Festinger's distribution-free '*d*' test (5).

It would clearly be too much to expect of the judgments based upon the Cleckley criteria that they should have perfectly separated the psychopathic sample into a "primary" species in Group I, and a neurotic or dissocial species in Group II. That the separation was reasonably good, however, is supported by the finding that Group II scored significantly higher than the normals on the Taylor scale, a great deal of evidence having accumulated (4, 7, 15) to indicate that this scale is primarily a measure of neurotic maladjustment or neuroticism rather than of anxiety level or anxiety reactivity *per se*. On the MMPI Anxiety Index, which like the Taylor scale is unquestionably polydimensional with a heavy loading on neuroticism, Group II again has the highest mean, with Group I again only slightly higher than Group III.

In contrast, the Anxiety scale, which was designed for this study and which is not loaded on neuroticism and only negligibly correlated with the Taylor scale or the AI, separated the groups in a different order. On this test, the primary types of Group I show the least anxiety reactivity, significantly less than the normals, with Group II falling in between but rather nearer to the Group III mean. This result appears to support hypothesis *b* of this study, that the subset of primary sociopaths show abnormally little manifest anxiety, i.e., anxiety reactivity to the real-life anxiety stimuli referred to in the questionnaire.

Both sociopathic groups scored significantly higher than the normals on the *Pd* scale of the MMPI, but this measure, which differentiates at the phenotypic or genus level, does not distinguish between the types or species of sociopathy represented in Groups I and II.

Schedule difficulties unfortunately led to a reduction in the number of *S*s to whom the avoidance learning test could be given. With nearly half of the total group, the available testing time was too short to cover all of the procedures; in such cases the avoidance test, requiring nearly an hour to give, was passed over. Even on the residual sample of 34 *S*s, however, rather clear-cut differences exist. As a crude, overall index of avoidance learning, the avoidance scores (shock errors divided by

unshocked errors) were averaged for all but the first of the 20 trials; this is the basis of the mean scores entered under "avoidance" in Table 1. The distribution was reversed to make high values represent greater avoidance of the shock. It is impossible, of course, to summarize adequately a complex learning process by a single numerical index of this sort, but in spite of these limitations, it is striking that Group I (primaries) shows the least avoidance as expected, Group II (neurotics) next, and Group III (normals) the most. The Group I versus Group III, and Group II versus Group III differences are significant by Festinger's *d*-test (5), and the actual distribution of scores shows the groups to be remarkably well separated (only 17 per cent overlap between Groups I and III). This result supports hypothesis *c* of this study, that the primary sociopath demonstrates defective avoidance learning.

Results of the GSR Conditioning Series

Of all the tests employed here, principal emphasis should be laid on GSR conditioning. The various difficulties attending the interpretation of GSR data are well known, but one fact stands out with relative certainty: given certain necessary conditions, if an *S* does *not* produce a GSR to a stimulus, one can be sure that he has not "reacted emotionally" to that stimulus.

The two numerical indices which were derived as alternative ways of representing in a single value the conditioning indicated by the GSR protocols (anticipatory GSR to the buzzer after several pairings with shock) have already been described. As shown in Table 1, the group means are in the expected order on both indicants, with Group I significantly lower than Group III on GSR Reactivity and GSR Conditioning (.05 level, *d*-test).

A somewhat more meaningful comparison is obtained by contrasting the reactivity by trials for the three groups. Group I shows the least GSR reaction to the CS in 14 out of the 16 double trials. Group II is significantly higher (.02 level) than Group I at the end of the extinction trials. The positions of Group II and Group III interchange during the series with Group II beginning to show greater reactivity during the extinction trials, suggesting a perseveration (failure of extinction) of the anxiety response in the neurotic group. This trend was tested for statistical reliability by correlating the differences between Group II and Group III with the ordinal position in the conditioning series at which the difference was taken. The quadrant sign test (24) shows this association to be significant at the .01 level. This result supports hypothesis *a* of this study, that the primary sociopath is defective in his ability to condition the anxiety response.

The generalization scores were leptokurtically distributed, the group differences being determined by a few deviant *S*s. Group II shows the highest mean generalization score, but the differences are not significant.

Summary

Forty-nine diagnosed psychopaths were divided into two groups according to the descriptive criteria of Cleckley. Fifteen normals served as controls. A battery of tests related to anxiety reactivity or anxiety conditionability were administered. As compared with normals, the Cleckley, or "primary" sociopaths, showed significantly less "anxiety" on a questionnaire device, less GSR reactivity to a "conditioned" stimulus associated with shock, and less avoidance of punished responses on a test of avoidance learning. The "neurotic" sociopaths scored significantly higher on the Taylor Anxiety Scale and on the Welsh Anxiety Index.

REFERENCES

1. Amer. Psychiatric Assoc. *Diagnostic and statistical manual: mental disorders*. Washington, D.C.: American Psychiatric Assn., 1952.
2. Bitterman, M. E., & Holtzman, W. H. Development of psychiatric screening of flying personnel. III. Conditioning and extinction of the GSR in relation to clinical evidence of anxiety. *USAF Sch. Aviat. Med.*, 1952, Proj. No. 21-37-002, Rep. No. 3, N. 232 p.
3. Cleckley, H. *The mask of sanity*. (2nd ed.) St. Louis: C. V. Mosby, 1950.
4. Eriksen, C. W., & Davids, A. The meaning and clinical validity of the Taylor Anxiety Scale and the hysteria-psychasthenia scales from the MMPI. *J. abnorm. soc. Psychol.*, 1955, **50**, 135–137.
5. Festinger, L. The significance of the difference between means without reference to the frequency distribution function., *Psychometrika*, 1945, **11**, 97–105.
6. Flanders, N. A. A circuit for the continuous measurement of palmar resistance. *Amer. J. Psychol.*, 1953, **66**, 295–299.
7. Franks, C. Conditioning and personality: a study

of normal and neurotic subjects. *J. abnorm. soc. Psychol.*, 1956, **52,** 143–150.

8. HAGGARD, E. A. Experimental studies in affective processes. II. On the quantification and evaluation of "measured" changes in skin resistance. *J. exp. Psychol.*, 1945, **33,** 46–56.
9. HAGGARD, E. A. On the application of analysis of variance to GSR data. I. The selection of an appropriate measure. *J. exp. Psychol.*, 1949, **39,** 378–392.
10. HATHAWAY, S. R. *Supplementary manual for the MMPI. Part I, The K scale and its use.* New York: Psychological Corp., 1946.
11. HEINEMAN, C. E. A forced choice form of the Taylor Anxiety Scale. *J. consult. Psychol.*, 1953, **17,** 447–454.
12. KARPMAN, B. Psychopathic types: the symptomatic and the ideopathic. *J. crim. Psychopathol.*, 1941, **3,** 112–124.
13. KARPMAN, B. The myth of the psychopathic personality. *Amer. J. Psychiat.*, 1948, **104,** 523–534.
14. KARPMAN, B. Conscience in the psychopath: another version. *Amer. J. Orthopsychiat.*, 1948, **18,** 455–491.
15. KERRICK, JEAN S. Some correlates of the Taylor Manifest Anxiety Scale. *J. abnorm. soc. Psychol.*, 1955, **50,** 75–77.
16. LACEY, O. L., & SIEGEL, P. S. An analysis of the unit of measurement of the galvanic skin responses. *J. exp. Psychol.*, 1949, **39,** 122–123.
17. LIPPMAN, H. S. Psychopathic behavior in infants and children: a critical survey of existing concepts. *Amer. J. Orthopsychiat.*, 1951, **21,** 227–231.
18. MAY, M. A. Experimentally acquired drives. *J. exp. Psychol.*, 1948, **38,** 66–77.
19. MILLER, N. E. Studies of fear as an acquirable drive. I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38,** 89–101.
20. MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 435–472.
21. MOWRER, O. H. A stimulus-response analysis of anxiety. *Psychol. Rev.*, 1939, **46,** 553–565.
22. MOWRER, O. H. Anxiety reduction and learning. *J. exp. Psychol.*, 1940, **27,** 497–516.
23. WELSH, G. S. An anxiety index and an internalization ratio for the MMPI. *J. consult. Psychol.*, 1952, **16,** 65–72.
24. WILCOXON, F. *Some rapid approximate statistical procedures.* New York: American Cyanamid Co., 1949.

Received July 5, 1956.

A STUDY OF AFFECTIVE RESPONSIVENESS IN A LIE-DETECTION SITUATION[1]

JACK BLOCK

Institute of Personality Assessment and Research, University of California, Berkeley

SOME years ago, H. E. Jones (9) found a relatively low relationship between the intensity of the galvanic skin response (GSR) and the amount of overt emotional expression when both were measured simultaneously in nursery-school children. Although stimuli which *on the average* elicited strong autonomic responses were also found to evoke a great deal of observable emotional expression, *for individual children* the correlation over situations between GSR activation and outward emotionality was quite low.[2] Since the measures employed were highly reliable, the low relationship within individuals between internal and overt manifestations of emotions suggested that stable individual differences exist in the patterning of emotional expression.

This provocative finding was repeated and extended by Jones in a further study (10). Here, one hundred adolescents were available who, in the course of participating in a longitudinal study, had experienced eleven test situations where GSR recordings were gathered. The 20 individuals who were highest in the average magnitude of their GSRs and the 20 who were lowest were selected, and the independently rated overt behavioral characteristics of the two groups compared.

The "high reactives" were rated as less assertive, less animated, less talkative, and less attention-seeking than the "low reactives," all of these differences occurring at the .01 level of significance. The high reactive group was also evaluated as more calm, more deliberative, more good-natured, more cooperative, and more responsible than the low reactives, who were judged as irritable, excitable, impulsive, and immature.

Jones suggested that the high reactives are introversive characters, while the low reactives represent the psychoanalytic concept of "impulse neurosis" (2). The process of socialization is the agent causing the child to shift from undifferentiated, externalizing modes of expression to complex, internalizing avenues of affect discharge.

It is important to note that Jones also observed the presence of two additional patterns of emotional expression. He identified a "generalizer" type, consisting of children who respond at both overt and internal levels simultaneously, and a "reciprocal" type, represented by children who sometimes respond overtly but with little GSR reactivity and who at other times evidence a highly reactive GSR but with little overt response.

The work by Jones has been summarized here because it is surprisingly little known and because it served as a springboard for the present experiment. His results have implications for conceptions of the nature and etiology of psychopathy and for a more general understanding of the place and process of affect in personality.

The present study aimed at an independent verification and extension of Jones's findings. It contrasts with the previous work in several ways: the subject sample used was older and probably more homogeneous with respect to intelligence, interests, and socioeconomic level; there was only one GSR recording session; the more stressful circumstance of lying was employed as the stimulus situation; and finally, objective test measures as well as ratings were available for comparison with GSR reactivity.

METHOD

Subjects and Experimental Procedure

Seventy male applicants to a medical school underwent an 18-hour assessment program at the Institute

[1] This study was supported in part by grants to the Institute of Personality Assessment and Research from the Rockefeller Foundation and the National Institute of Mental Health of the National Institutes of Health, U. S. Public Health Service. Assessment is a group endeavor, and the writer is pleased to record his debt to the staff of the Institute who participated in the program and who helped provide the data reported here. The writer is particularly grateful to Samuel Fillenbaum for his advice and assistance.

An abbreviated version of this paper was read at the meeting of the American Psychological Association at San Francisco, California, September, 1955.

[2] The correlation over the 22 situations between the group's average GSR and the average rating on overtness of response was .83 and .80 in the initial and retest experiments. For individual children, the correlation averaged .20 and .25 in the two separate tests.

of Personality Assessment and Research. In age they ranged from 19 to 32 years with a mean at 22. Intellectually, the group was at a very superior level. Their socioeconomic level was clearly middle class.

Each of the subjects (*Ss*) was observed in a wide variety of personality revealing situations, ranging from charades and stress interviews to standardized psychodramas and dinnertime conversation. One of the procedures each *S* experienced was the lie-detection situation. At his scheduled time, *S* entered the experimental room where he was greeted by the experimenter (*E*) and seated before the imposing apparatus. A pneumograph was attached around *S*'s chest, his left thumb was inserted into a pulse pressure recording device, and electrodes were attached to his right hand. Only the GSR circuit of the apparatus was in fact operative during the experiment, the respiration and pulse transducers being employed for their dramatic effect only.[3]

The *E* conducted a short interview, inquiring about the origins of *S*'s medical interests, his professional aspirations, the nature of his father's occupation, and so on, while an assistant adjusted the apparatus. *S* was then informed that the apparatus was a lie detector and that his skill as a liar was to be investigated.

The *S* selected a number ranging from 1 to 5 without informing *E* of his choice. The *E* then queried *S* a number of times over each possibility, employing a fixed but unsystematic order of interrogation. To each question, *S* was instructed to say "no," even where denial meant a lie. Questioning was varied in phrasing and accentuation so as to maintain the interpersonal aspects of the essentially standard stimulus situation. At the end of the series of 30 questions, *S* informed *E* as to his actual choice. In two additional replications, *S* selected an alternative (one of six months, one of six colors) and, again, *E* repeatedly interrogated him as to his choice. Again, *S* denied all possibilities, including the one he had in fact chosen. The temporal relations of stimuli and responses were recorded by noting *E*'s questions on *S*'s GSR record.

It was found that some *Ss* displayed a diffuse and unselective GSR responsivity and some were relatively unreactive; most were selectively responsive, the GSR tending to appear only after the lie. Over the three trials, an accommodation effect was usually observed as the *S* adapted to the situation.

Two raters judged the GSR records for amount of reactivity and selected the 20 most highly reactive individuals and the 20 least reactive individuals. Reactivity was judged primarily on the basis of the number of GSRs the subject manifested without further quantification of the size or duration of the response. Ratings were employed rather than a simple count because of the necessity to evaluate occasional movement and instrument artifacts. Classification of *Ss* into the categories was highly reliable, 36 of the 40 criterion cases being selected by both raters. The disagreements were resolved by mutual consent of the raters. Categorization of *Ss* into the two GSR criterion groups occurred before the examination of any other data. The results to be reported in the next section are based upon the comparison of the two groups, Reactors and Nonreactors, on a number of measures.

[3] Although there is some controversy among practitioners as to the suitability of the GSR as a lie detector under actual field conditions, there appears to be agreement as to its validity in experimental contexts (8, 11).

RESULTS

Intellectual Differences

Intelligence is a powerful determiner of behavior, and its role in any particular context requires evaluation. Accordingly, the Reactors and Nonreactors were compared on measures defining various aspects of intellectual functioning. These included the Medical College Admissions test, grade-point average, a college vocabulary test (5), the Gottschaldt test (3), and a measure of originality based upon three of the Guilford tests of originality (6).

On none of these measures did the differences between the two groups approach significance. There is no evidence that such differences as may exist between the two groups can be ascribed to intellectual factors as these are ordinarily understood and measured.

Differences on Rating Measures

At the end of assessment, each *S* was evaluated by each staff psychologist. One of the means employed for codifying staff impressions of each *S* was the Q-sort method of personality description (13); another was the Gough Adjective Check List (4).

A comprehensive 115-item Q set was used which included items oriented toward highly inferential features of the personality system as well as toward the more readily observable aspects of the individual.[4] Five psychologists, none of whom had observed the *S*'s behavior in the lie-detection situation, Q sorted each *S*. These data, combined into a composite by equally weighting the ratings of each judge, codified the impressions and evaluations of these staff members with respect to the subject. Because the psychologist raters had no information as to the behavior of the *S* in the test situation, the Q sort data provide a totally independent source of information as to the personality characteristics of each *S*.

The comparison of the Q ratings for Reactors with the ratings for Nonreactors revealed twelve items differentiating at or beyond the

[4] A copy of the Q set employed is available from the writer.

.05 level of significance. These are listed in Table 1.

A separate group of seven staff psychologists employed the Adjective Check List in their characterization of each *S*. These data, when summed across raters and analyzed, served as a cross check on and amplification of the Q findings. The adjectives differentially characterizing the Reactors and Nonreactors at or beyond the .05 level of significance are presented in Table 2.

It is clear from the nature of the differentiating Q items and adjectives that the two groups differ along some version of the psychological dimension suggested by Jones. The Reactors are perceived as withdrawing, worrying individuals who turn their anxieties toward internal routes of expression. The Nonreactors are evaluated as independent, aggressively direct, and relatively nonconforming, all of these being visible rather than inward or covert expressions of impulses.

TABLE 1

Q ITEMS DISTINGUISHING BETWEEN THE REACTOR AND NONREACTOR GROUPS

Characteristic of Reactors:	
10. His anxiety and tension find outlet in bodily symptoms and dysfunction; anxiety and tension are converted into somatic symptoms	.05
11. Is protective of those close to him	.05
14. Is readily dominated by others; submissive	.05
19. Seeks reassurance from others	.05
30. Gives up and withdraws where possible in the face of frustration and adversity	.05
40. Is vulnerable to real or fancied threat; generally fearful; is a worrier	.05
70. Behaves in an ethically consistent manner	.05
74. Lets other people take advantage of him; allows exploitation	.05
Characteristic of Nonreactors:	
94. Expresses his hostilities directly	.01
96. Values his own independence and autonomy	.01
1. Critical, not easily impressed, skeptical	.05
62. Tends to be rebellious and non-conforming	.05

Inventory Items Differentiating the Groups

A number of personality inventory items (from the Minnesota Multiphasic Personality Inventory and the California Psychological Inventory) distinguished between the two groups. Summarizing their import, it would appear that the Reactors introspect relatively often, are sometimes radically subject to their moods and impulses, and are very much concerned with the appropriateness of their behavior in social contexts. In contrast, the Nonreactors indicate an absence of rumination and a self-control of their behavior. Also expressed is a preference for autonomy and a difficulty with representatives of authority.

Some Test Measures Differentiating the Groups

All *S*s experienced the rod-and-frame procedure developed by Witkin (14). Here the *S* is placed in a totally dark room before a luminescent square tilted from the vertical. He revolves, via a suitable control, a rod centered within the frame until he perceives the rod as at the true vertical (or horizontal). In the blackness of the experimental room, the only external cue to verticality is the misleading frame. An alternative basis for judgment is available from internal, proprioceptive gravitational cues. In the present study, the Reactors proved to be more affected by the external frame of reference than were the Nonreactors ($p < .05$). That is, the Reactors were more externally oriented and more open to outside influence in the presence of uncertainty than were the Nonreactors, who lent greater credence to their proprioceptive cues.

A rather curious result emerges from a comparison of the behavior of the two groups in the Crutchfield independence-of-judgment experimental procedure (1). Here the subject is exposed to a group consensus about a problem which differs from what the *S* perceives as the correct or preferred solution. The extent to which *S* conforms to the group opinion in his manifested choice provides a measure of independence found by Crutchfield to be richly related to character structure.

TABLE 2

ADJECTIVES DISTINGUISHING BETWEEN THE REACTOR AND NONREACTOR GROUPS
(.05 level of significance)

Characteristic of Reactors:	
cautious	idealistic
dependent	mannerly
dreamy	suggestible
Characteristic of Nonreactors:	
clever	leisurely
cool	opportunistic
evasive	practical
independent	realistic
ingenious	

In the present analysis, Reactors and Nonreactors did not differ in amount of independence as expressed by their choice behavior. But subjects also expressed their subjective confidence in their judgments. The Reactors, although actually conforming no more than Nonreactors, proved to be less confident about their course of action ($p < .001$). Independence (or conformity) in the Reactor thus appears to have a different psychological base than in the Nonreactor. It might even be said that manifesting independence in the face of uncertainty and doubts, as did the Reactors, required more courage than equivalent behavior where subjective doubt does not prevail, as among the Nonreactors.

Although the classification into Reactors and Nonreactors was based solely upon the evaluation of the GSR, some evidence for the physiological generality of reactivity may be adduced from a subsequent finding. Following a stressful psychodramatic situation (7), it was found that the systolic blood pressure of Reactors was raised significantly more than that of the Nonreactors ($p < .01$).

DISCUSSION

The present results, taken together with those of Jones and others, provide reasonable evidence that electrodermal responsiveness is related to personality structure and behavior. While in both studies, the number of significant relationships is small considering the multiple tests of significance, the consonance of both sets of data argues for their psychological stability. The placing of these data within a conceptual frame of reference is another matter.

An interpretation of the findings in terms of the notion of externalization-internalization simply does not serve. The externalizing-internalizing dimension has been only informally conceptualized, but it appears to be concerned with whether need tensions are discharged via external, action modes of expression or whether needs are routed internally into cognitive and visceral channels of discharge. Implicit here is the assumption that there exists a reciprocal relationship between overt motor response and cerebral or autonomic activity. That is, overt motor response *displaces* internal activity and *vice versa*. Jone's finding, supported by Seymour (12), that there are "generalizers," individuals whose expression of need states is simultaneously overt and covert, upsets such a notion. Observations of a culture like the Italian, where emotion is vigorously displayed but also apparently internally experienced, also weigh against the reciprocal theory of externalization-internalization. Clearly, an additional dimension or a reformulation of the concept is required in order to encompass the various patterns of emotional expression which have been noted.

It is suggested here that an important distinction is being obscured in the definition of externalization-internalization by opposing overt motor response on the one hand against cognitive *and* visceral channels of discharge on the other. The italicization of *and* is important, for it is precisely this conjunction which may be inappropriate. From an instrumental point of view, cognitive activity *and* motor behavior are both means to drive reduction, whereas autonomic, visceral reaction serves no such function. Autonomic activity appears instead to be an expression of an affective state and does not in and of itself bring about a significant reduction of the disturbing emotional charge. Tension reduction is brought about only by appropriate cognition, fantasy, or motor behavior, not by autonomic arousal.

If this reformulation appears a tenable one, then externalizing-internalizing can be redefined as a way of characterizing the direction or orientation of an individual's response to the impelling affective disequilibrium. GSR lability, from this point of view, is simply a convenient index of the tensional state, of whether an individual has anxiety or affect in a given situation or whether he does not have anxiety or emotion in that situation. It has no implication, however, for the direction of the response precipitated by the emotion. Since affective responsiveness may eventuate in motor as well as cognitive behavior, the dilemma previously posed by the "generalizer" no longer exists.

The dimension distinguishing between the Reactors and Nonreactors in the present study, may be the dimension of affective responsiveness without regard for just how the individual copes with emotional states. There has been little direct investigation of affective responsiveness although it is known that there are great differences in the extent to which individuals experience affect and have their self-

percept brought forward into consciousness as a conditioner of behavior. Some individuals experience too much affect, e.g., anxiety neurotics, and are thereby incapacitated. But others do not experience sufficient emotion, sufficient in relation to a normative criterion, and therefore do not react on the basis of an implicit but culturally shared set of behavioral premises. Because they operate from a different premise system, these individuals are more likely to emit behavior which is inconceivable to the affectively responsive observer. Such people include the primary psychopath, certain forms or stages of the schizophrenic reaction and, in general, those individuals identified by psychiatrists as possessing "flattened affect."

SUMMARY

Seventy *Ss*, during the course of an assessment program, experienced a lie-detection situation where GSR reactions were recorded. The twenty most reactive subjects and the twenty least reactive subjects were selected and compared on a variety of measures. Reactors appeared to be more dependent, dreamy, idealistic, and suggestible; Nonreactors were evaluated as relatively cool, evasive, opportunistic, and independent. The findings were related to the previous work of Jones, and a reformulation of the notion of externalization-internalization was offered.

REFERENCES

1. CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychologist*, 1955, **10**, 191–198.
2. FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
3. GOTTSCHALDT, K. Uber den einfluss der erfahrung auf die wahrnehmung von figuren. *Psych. Forsch.*, 1926, **8**, 261–317; 1929, **12**, 1–87.
4. GOUGH, H. G. *The adjective check list*. Berkeley: Univer. California Printing Department, 1952.
5. GOUGH, H. G., & SAMPSON, H. *The college vocabulary test*. (Form A). Berkeley: Univer. of California Inst. of Personality Assessment and Research, 1954.
6. GUILFORD, J. P., WILSON, R. C., CHRISTENSEN, P. R., & LEWIS, D. J. A factor-analytic study of creative thinking: I. Hypotheses and descriptions of tests. *Reports from the Psychol. Lab.*, No. 3, Los Angeles: Univer. of Southern California, 1951.
7. HARRIS, R. E., SOKOLOW, M., CARPENTER, L. G., FREEDMAN, M., & HUNT, S. P. Response to psychologic stress in persons who are potentially hypertensive. *Circulation*, 1953, **7**, 874–879.
8. INBAU, F. E. *Lie detection and criminal interrogation*. (2nd ed.) Baltimore: Williams & Wilkins, 1948.
9. JONES, H. E. The galvanic skin response as related to overt emotional expression. *Amer. J. Psychol.*, 1935, **47**, 241–251.
10. JONES, H. E. The study of patterns of emotional expression. In M. L. Reymert (Ed.), *Feelings and emotions: the Mooseheart Symposium*. New York: McGraw-Hill, 1950.
11. MACNITT, R. D. In defense of the electrodermal response and cardiac amplitude as measures of deception. *J. crim. Law and Criminol.*, 1942, **33**, 266–275.
12. SEYMOUR, R. B. Personality correlates of electrodermal resistance and response. Unpublished doctor's dissertation, Univer. of California, 1950.
13. STEPHENSON, W. *The study of behavior*. Chicago: Univer. of Chicago Press, 1953.
14. WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception*. New York: Harper, 1954.

Received July 24, 1956.

THE IMMEDIATE AND DELAYED EFFECT OF FAILURE AS A FUNCTION OF TASK COMPLEXITY AND PERSONALIZATION OF FAILURE

CHARLES B. TRUAX AND BARCLAY MARTIN

University of Wisconsin

Previous research has suggested various factors as being important in determining the effects of experimentally induced stress upon performance. This study is concerned with the manipulation of several of these factors within the framework of one experimental design. There is ample evidence that individual differences in general anxiety level, whether measured by paper-and-pencil tests or assessed by psychiatric diagnosis, are related to performance under stress conditions (1, 3, 4, 5, 9, 14, 15). The typical finding in this respect is an interaction of anxiety level by stress, so that high anxious subjects (*Ss*) tend to do less well under stress than under neutral conditions, and low anxious *Ss* tend to perform relatively better under stress than under neutral conditions.

This finding, however, cannot be accepted without qualification, and it was essentially the purpose of this study to investigate some of the factors which may affect this interaction of anxiety level and stress. One such factor is the level of complexity of the task being performed or learned. Studies have generally indicated that the more complex the task, the more poorly the high anxiety *Ss* do in relation to the low anxiety *Ss* (2, 7, 10, 12, 13). Thus, in the present experiment two levels of complexity have been introduced.

Another factor which may influence the effects of failure (as one form of stress) on performance is the length of the time interval interposed between the failure and the postfailure performance. This variable has usually been described in terms of the immediate versus delayed effects of failure, and previous findings are somewhat conflicting. Russell (8) found failure conditions to be associated with poorer retention scores than nonfailure conditions under immediate, but not delayed (24 hours), recall conditions. The task used was a list of 12 nonsense syllables learned by the serial anticipation method. Farber, Russell, and Andreas (1), using a paired-associate verbal learning task, found that failure adversely affected performance on immediate tests of retention; but on delayed tests, the failed group tended to perform better than the controls. Only the decrement under the immediate testing conditions was significantly different between the failed and nonfailed groups however.

On the other hand, Zeller (16, 17) found both an immediate and delayed (72 hours) effect of failure relative to nonfailure conditions, using both a recall and a relearning score based on a paired-associate task. One clue suggesting an explanation for the discrepancy between Zeller's study and the others mentioned above may be found in the nature of the failure comments given. Zeller's failure remarks were without a doubt extremely severe and were to a considerable extent personally insulting to *S*. The comments in the Russell and the Farber, Russell, and Andreas studies, while without question indicating to *S* that he failed, were clearly much less severe and personalized. It may be, then, that delayed effects of failure become apparent only under rather strongly personalized conditions. In other words, failure experiences which are explicitly made to reflect upon the general competence and integrity of *S* may have longer lasting effects than those which imply only failure on the immediate performance.

To explore these possibilities, immediate and delayed postfailure testing, mild and severe personalization of failure, two degrees of task complexity, and subgroups divided at their medians on the basis of either the Taylor Manifest Anxiety Scale or the Stroop Interference Test were all incorporated as variables within the experimental design. The Stroop test (11) was included because it seemed likely that it might give some measure of a person's general susceptibility to interfering response tendencies, a trait that Waterhouse and Child (14) have suggested as being closely related to quality of performance under stress.

Method

The *Ss* were 128 female students in the introductory psychology course at the University of Wisconsin. Sixteen *Ss* were randomly assigned to each of 8 treatment

groups. Since the two different tasks involved different dependent variables, the results had to be analyzed separately; accordingly, the study can be described as consisting of two separate two-by-two factorial designs. Thus, for the simple task condition, there were two levels of personalization (mild and severe) of failure orthogonal to two intervals between pre- and postfailure testing (20 seconds and 24 hours). In the complex task condition, a similar design was used. The individual difference measure was introduced by dividing the subgroups at the median on either the Taylor Manifest Anxiety Scale or the Stroop Interference Test, adding an additional dimension to the factorial design. Since Taylor scores were not obtained on several *Ss*, it was necessary to reduce all subgroups to an *N* of 14 in order to retain cell proportionality for this measure. *N*s of 16 were used for the Stroop test. The Taylor Manifest Anxiety Scale was administered to the *S*s as a group in their regular class period before the individual testing session.

The Stroop Interference Test was given at the beginning of the individual session. This test consists of three parts: Card A, a page on which only the four color names of red, blue, yellow, and green are printed in black ink in a random arrangement of eleven rows of ten words each; Card B, a page on which the same colors appear in ¾ inch colored dots, arranged randomly in the same format as Card A; and Card C, a page on which the same colors are printed in random order, with each color name printed randomly in a color other than the one it designates. The *S*'s task was (*a*) first to read as fast as possible the color names on Card A, (*b*) to name as fast as possible the colors of the color dots on Card B, and (*c*) to name as fast as possible the colors on Card C in which the words were inked, ignoring any interference produced by the "wrong" names. The score used was the ratio of the time to finish Card B to the time to finish Card C.

The simple task. The simple task consisted of *S*'s adding 78 pairs of one-digit numbers. The score was the time in seconds to complete the problems. Uncorrected errors on this extremely simple task were not frequent enough to affect the time score. The same task was used in the postfailure testing.

The complex task. The complex task consisted of the Arithmetic subtests of the Wechsler-Bellevue Intelligence Scale, Forms I and II. One form was presented under prefailure, and the second form was given under postfailure conditions with order counterbalanced within each of the 8 major experimental groups. The weighted score derived from the Wechsler-Bellevue standardization norms, involving both time and accuracy measures, was used.

The failure inducing task. Failure comments were given for a totally different task so that all *S*s would experience failure in a common situation. This task was an extended version of the Wechsler-Bellevue Digit Symbol Substitution Test. It was presented on a dittoed sheet with the label, "Wisconsin Digit Symbol Test of Intellectual Ability, College Form A," prominently displayed across the upper third of the page. The time allotted for the completion of this task was varied with *S*'s speed to insure that *S* would complete only three-fourths of the test.

Procedure and instructions. The following instructions were given initially for the Stroop Interference Test:

> We are in the process of obtaining normative data on a test which we hope will be a good measure of Reading Ability. This test is in three parts, and we'll start off with the first part now.

The Stroop test was then administered, after which *S* was told,

> As you may know, in standardizing a test you have to get other measures on the subjects in order to check for a normal distribution. For example, if all of our subjects were exceptionally intelligent or quick reacting, the Reading Test wouldn't be very good for estimating the ability of the average college student. So we have some other tests for you to take.

The simple or complex task was then administered, depending on the treatment group to which *S* was assigned. After this task, the following "explanation" was given of the failure inducing task:

> Reading ability, as you probably know, is related to IQ, so to standardize the Reading Test for a college population, we will have to get a measure of the IQ of the subjects used in this study. We will use the Wisconsin Digit Symbol Test. It's one of the shortest IQ tests available, but it's still very reliable. The average college student will get about 114; the brightest student so far in this study has got 129 items right.
>
> Don't expect to finish the test entirely in the two minutes you will have—just do the best you can.

After the failure inducing task was administered, experimenter (*E*) laid a template over the test and pretended to score it. The following comments (which had previously been memorized by *E*) were then made in the "severe personalization" group:

> Damn! (*E* frowns and shakes head). Ninety-two. You *really* goofed this test . . . . . (*S*'s first name). You're well below the norm of 114. (Pause.) You having much trouble keeping up with your classes? Well, either you were just goofing up this experiment . . . . . (*S*'s first name), or something's wrong with you. A score of 92! Damn!!!

The following comments (which had previously been memorized by *E*) were made in the "mild personalization" group:

> Ninety-two. (*E* pauses and continues in a moderate tone.) The score is 92. (Pause.) This is below the norm and not very good. (Pause.) A score of 92!

E paced the mild and severe personalization remarks so that the same amount of time was taken in giving both, about 17–18 seconds.

After the induction of failure, half the *S*s were retested immediately on the performance task (either simple or complex), and an appointment was made for the other half to meet for a second session 24 hours later. All *S*s were given the following instructions before the administration of the postfailure test:

> Well, let's do this arithmetic (addition) test over again. Remember, I'll be timing you, so work as quickly as possible without making mistakes.

At the conclusion of the experiment, each *S* was told the purpose of the study and allowed to vent her feelings about the failure experience.

RESULTS

Simple task. There was a negative correlation of −.68 between scores on the simple task

under the prefailure condition and the change score from pre- to postfailure. Also, by chance apparently, the mild and severe personalization groups differed greatly in their prefailure scores. Accordingly, analysis of covariance was used, making it possible to analyze only that variance in the change scores that was uncorrelated with the prefailure scores. The variance within the different treatment groups did not differ significantly as indicated by Bartlett's test for homogeneity of variance.

It should be mentioned, before these analyses are considered, that a previously performed analysis of variance with repeated measures indicated a significant tendency ($p < .001$) for all *Ss* to improve their performance on this simple task after failure. This general tendency would not appear, of course, in the analysis of covariance in which change scores were used; but it indicates that any found effects will represent differential tendencies to improve from pre- to postfailure conditions.

The Stroop test did not prove to be significantly related to change in performance on either the simple or complex tasks, and results with respect to this measure were not further considered.

Table 1 presents the analysis of covariance with the subgroups divided into high and low Taylor scale scores. The main effect of the immediate-delay variable was significant in the direction of greater improvement in performance under the delayed postfailure testing. Individual differences in anxiety were also significantly related to change in performance. High anxious *Ss* showed greater improvement ($p < .01$) than low anxious *Ss*, and there was an interaction of anxiety with mild-severe personalization ($p < .001$), suggesting that high anxious *Ss* improved *less* under severe personalization than under mild, whereas low anxious *Ss* improved *more* under severe than under mild personalization.

TABLE 1

ANALYSIS OF COVARIANCE FOR THE SIMPLE TASK USING THE TAYLOR SCALE

Source	*df*	Mean Square	*F*
Taylor scale (A)	1	110.94	7.25**
Personalization (B)	1	1.87	—
Immediate-Delay (C)	1	76.80	5.02*
A × B	1	203.44	13.24***
A × C	1	41.89	2.74
B × C	1	15.47	1.01
A × B × C	1	7.22	—
Within groups	47	15.31	—

* $p < .05$.
** $p < .01$.
*** $p < .001$.

TABLE 2

ANALYSIS OF COVARIANCE FOR THE COMPLEX TASK USING THE TAYLOR SCALE

Source	*df*	Mean Square	*F*
Taylor scale (A)	1	9.05	2.86
Personalization (B)	1	0.01	—
Immediate-Delay (C)	1	30.93	9.76**
A × B	1	0.49	—
A × C	1	0.05	—
B × C	1	1.87	—
A × B × C	1	12.09	3.82
Within groups	47	3.17	

** $p < .01$.

This interaction was further investigated by performing an analysis of covariance upon the high and low groups separately. The low anxiety group was found to perform significantly better under severe than mild personalization ($p < .05$), but there was no evidence that the high anxiety group performed better under mild than severe personalization. The interaction effect was apparently confined to the lower portion of the anxiety distribution.

Complex task. A repeated measures analysis of variance revealed no significant tendency for *Ss* in general to change in performance from pre- to postfailure on the complex task. Initial scores correlated with change scores −.54, so analysis of covariance was also used here. Table 2 presents the analysis of covariance for subgroups divided on the basis of the Taylor scale. The immediate-delay variable was significant ($p < .01$), with *Ss* again showing more improvement in performance under the delayed postfailure testing. No other variables were significant.

It is sometimes reported that variability increases as a function of failure (6). The variability in the prefailure situation was compared to that in the postfailure situation for both simple and complex tasks. Changes were small and insignificant in both cases; the *F* values were, respectively, 1.19 and 1.02.

The differences in results between the simple and complex tasks do not appear to be due to differences in the extremeness of groups based on the Taylor scale. Table 3 presents the mean Taylor scale scores for the groups and their

TABLE 3
MEANS AND PERCENTILE RANKS OF GROUPS ON TAYLOR SCALE

Group	Simple Task		Complex Task	
	Mean	P.R.	Mean	P.R.
High anxiety	23.22	86	20.95	80
Low anxiety	9.72	32	10.00	34

percentile rank as estimated from the University of Iowa normative data.

DISCUSSION

No main or interactive effects were found when the Stroop Interference Test was used as the individual difference measure. These negative results suggest that anxiety is either not fruitfully conceptualized in terms of a general susceptibility to interfering responses as Waterhouse and Child (14) suggest or that the Stroop test is not a valid measure of this trait.

When the Taylor scale is used to assess individual differences, significant relationships were found with pre- to postfailure performance changes for the simple but not for the complex task. This suggests that the Wechsler Arithmetic Subscale may be relatively less sensitive to the effects of failure than the simple additions task, although there was a tendency for performance on the Arithmetic subscale to be affected by the immediate-delay variable. Both tasks involve relatively well learned past habits, and this common aspect of both tasks is probably not particularly susceptible to the effects of failure and general anxiety level. The simple additions test, however, is much more predominantly a speed test, and it may be that its relationship to the failure and anxiety variables is essentially a function of the speed factor, which seems much less important in the Wechsler problems.

The significant effects of the immediate-delay variable for both the simple task and the complex task suggest that failure has a greater effect after a 24-hour delay than under immediate retesting conditions. Nevertheless, this interpretation requires considerable caution since, first, it may hold only for conditions in which performance improves after failure, and second, no control for practice effects was available in this experiment. With regard to the latter consideration, however, the contention would have to be that practice has a more beneficial effect after a 24-hour delay than upon immediate retesting, which seems possible but rather improbable.

The results of this study are in some respects consistent with those of Farber, Russell, and Andreas (1), who found that a failed group did worse than a nonfailed control group under immediate retesting conditions, but tended to do better than the control group under delayed retesting. Russell (8) also found the negative effects of failure to be greater in the immediate than in the delayed situation. With the present results, these findings suggest that the detrimental effects of failure are more pronounced in the immediate than in the delayed situation, whereas the facilitating effects of failure are more pronounced in the delayed than the immediate situation.

For the simple task, it was found that high anxious *Ss* showed greater improvement after failure than low anxious *Ss*. This result appears to contradict the usual findings. Williams (16), employing a similar task of simple addition problems, but using the Waterhouse and Child Interfering Tendency Questionnaire to select high and low groups, found an interaction in the opposite direction ($p < .10$), with high scorers performing about the same after nonfailure as after failure conditions. Waterhouse and Child (14) report similar interactions with somewhat different kinds of tasks.

However, if one introduces the variable of mild-severe personalization of failure, the results of the present study no longer seem so contradictory of previous findings. The low anxious *Ss* show more improvement under severe than mild personalization of failure, whereas the high anxious *Ss* do not show this differential improvement. The resulting interaction of anxiety with mild-severe personalization therefore seems to represent a psychologically meaningful process. Thus, it looks as though the performance of high anxiety *Ss* on this relatively simple task is not adversely affected by either mild or severe failure. On the contrary, their performance is comparably facilitated by the additional anxiety or motivation under both conditions. The low anxiety *Ss*, however, show a greater improvement under severe than mild personalization. It may be that high anxiety *Ss* are approaching an optimum state of tension in the mild conditions and, accordingly, further stress does not enhance performance; but the low anxious *Ss*,

having not yet reached the optimum level of tension, continue to improve in performance under severe personalization. This finding is consistent with the general hypothesis that adequacy of performance as a function of anxiety is a curvilinear function of an inverted U-shape type. The fact that the high anxiety group has apparently not gone beyond the optimum state of tension to the point where performance begins to be disrupted may be due to the extremely simple nature of the one-digit additions task. This interaction of anxiety level with degree of personalization also indicates the importance of reporting in detail the operations defining the failure in any given piece of research, since this variable seems to affect performance significantly.

SUMMARY

This experiment was designed to study the effects of several variables upon changes in performance from pre- to postfailure conditions. All *Ss* were female college students. For one group of *Ss* performance was measured on a simple task, one-digit addition problems; for another group of *Ss*, performance was measured on a more complex task, the Arithmetic subscale of the Wechsler-Bellevue. A factorial design was used for each of these groups with the independent variables being mild-severe personalization of failure, immediate-delay of postfailure retesting, and high and low groups based on either the Stroop Interference Test or the Taylor scale.

All *Ss* tended to improve in performance after failure on the simple task. High anxious *Ss* tended to improve more than low anxious *Ss*. The interaction of anxiety with mild-severe personalization was significant with low anxious *Ss* tending to improve more under severe than mild personalization, and high anxious *Ss* improving about equally under both conditions. The immediate-delay variable was significant with *Ss* showing greater improvement after the 24-hour delay.

The only relationship found, when performance on the complex task was used as the dependent variable, was a tendency for the *Ss* to improve more after the 24-hour delay than upon immediate retesting.

The findings in regard to the immediate delay variable suggest the tentative hypothesis that the facilitative effects of failure are greater after a delay, but the detrimental effects are greater when retesting is done immediately.

REFERENCES

1. FARBER, I. E., RUSSELL, W. A., & ANDREAS, B. G. The effect of failure upon performance in verbal and motor learning situations: A review of work done at the State Univer. of Iowa. Paper read at Midwest. Psychol. Ass., Chicago, April, 1949.
2. FARBER, I. E., & SPENCE, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.
3. KAYE, D., KIRSCHNER, P., & MANDLER, G. The effects of test anxiety on memory span in a group test situation. *J. consult. Psychol.*, 1953, **17**, 265–266.
4. LUCAS, J. D. The interactive effects of anxiety, failure, and intra-serial duplication, *Amer. J. Psychol.*, 1952, **65**, 59–66.
5. MALMO, R. S., & AMSEL, A. Anxiety-produced interference in serial rote learning with observations on rote learning after partial frontal lobectomy. *J. exp. Psychol.*, 1948, **38**, 89–101.
6. MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.
7. MONTAGUE, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.
8. RUSSELL, W. A. Retention of verbal material as a function of motivation instructions and experimentally induced failure. *J. exp. Psychol.*, 1952, **43**, 207–216.
9. SARASON, S. B., MANDLER, G., & CRAIGHILL, P. G. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 561–565.
10. SPENCE, K. W. Current interpretations of learning data and some recent developments in stimulus-response theory. In The Kentucky symposium, *Learning theory, personality theory, and clinical research*. New York: Wiley, 1953.
11. STROOP, J. R. Studies of interference in serial verbal reactions. *J. exp. Psychol.*, 1935, **18**, 643–661.
12. TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.
13. TAYLOR, JANET A., & CHAPMAN, J. P. Paired associate learning as related to anxiety. *Amer. J. Psychol.*, 1955, **68**, 671.
14. WATERHOUSE, I. K., & CHILD, I. L. Frustration and the quality of performance: III. An experimental study. *J. Pers.*, 1954, **21**, 298–311.
15. WILLIAMS, J. E. Mode of failure, interference tendencies, and achievement imagery. *J. abnorm. soc. Psychol.*, 1955, **51**, 573–580.
16. ZELLER, A. F. An experimental analogue of repression: II. The effect of individual failure and success on memory measured by relearning. *J. exp. Psychol.*, 1950, **40**, 411–422.
17. ZELLER, A. F. An experimental analogue of repression: III. The effect of induced failure and success on memory measured by recall. *J. exp. Psychol.*, 1951, **42**, 32–38.

Received July 30, 1956.

USES OF COMMON OBJECTS AS INDICATORS OF CULTURAL ORIENTATIONS[1]

WAYNE DENNIS

Brooklyn College

If children are asked with regard to each of several objects, "What is it for?" one finds that children's concepts of the functions of identical things vary from society to society. Those functions of common objects which are frequently mentioned by the children of a social group, it is here proposed, are indicative of the activities, interests, and values that prevail in that society. That is, it is held that children value things for the uses to which they may be put, and, conversely, that frequently-mentioned uses are indicative of values. For example, if children in one society often mention boys and girls, cats and dogs, trees and sand, as functioning chiefly in play whereas the children of another group mention them more often in connection with work, these facts suggest that there is a difference between the two groups in regard to work-play emphases. The present paper illustrates the application of this method, The Uses Test, with three groups of children, an American group, a Lebanese group, and a Sudanese group.

If the objects used in questioning children are universally present, this test procedure can be employed in all cultures. Techniques for the study of child behavior that are universally applicable are very small in number. The method employed in the present study is offered as a useful addition to the small stock of tools available for cross-cultural research. It is believed to be particularly useful as an exploratory technique in societies whose patterns of child behavior are but slightly known to the investigator.

[1] The research here reported was done during 1955–56 when the author was a visiting professor at the American University of Beirut, Lebanon. Expenses of the study were defrayed by a grant of the Rockefeller Brothers Fund to the American University of Beirut. The writer wishes to express his gratitude to Mrs. Adele Hamdan Taky Din, Miss Leila Biksmati, Mrs. Marie Therese Broussalian, Mrs. Yvonne Sayegh and Miss Aimee Sabbagh who served as research assistants, and to the principals, teachers, and pupils whose cooperation made the study possible. He is also greatly indebted to staff members of the departments of psychology and education, and to members of the child development seminar at the American University of Beirut, who gave invaluable assistance and counsel throughout the course of the investigation.

Method

The question employed is the simple and probably universal one, "What is ——— for?". Naturally, this question may be asked in regard to any plant or animal, any class of persons, or any part of the body.

The question, "What is ——— for?," was chosen because numerous studies indicate that the young child readily thinks in terms of use. He spontaneously asks what things are for, he defines objects in terms of their use, and he seems to presume that every object in the world has a use. No difficulties have been encountered in obtaining answers to the question from hundreds of children between ages five and ten in a variety of cultures. In all likelihood it can be used above and below the age levels represented in this study. The technique is not limited to the particular word list employed here. The objects concerning which children are questioned will vary with the interests and hypotheses of the investigator.

In the present study, the following words were used in the order given: the mouth, the hands, a mother, a father, a boy, a girl, trees, wood, dogs, cats, birds, stones, the sun, sand, rain, and gold. In the case of the American and Lebanese subjects, data were gathered by individual interview. The Sudanese subjects were instructed orally as a group, and they were requested to write their answers.

Although the question put to the child did not ask for more than one use for each object, some children gave more than one. In the present study, responses beyond the first were not included in the analysis of the data, unless specific reference is made to answers beyond the first one.

Subjects

American. These were American children attending the American Community School in Beirut, Lebanon. The majority of these children were born in the United States and spent their pre-school years there. They were in Lebanon in 1955–56 because their parents were temporary residents of that country. Most of the fathers were employed by the University, or American business firms. Nearly all could be classified as belonging to the American middle class. All children between ages 5.0 and 11.0 were tested. These totaled 120 children, approximately equally representing the two sexes and the various ages.

Lebanese. This group consisted of an analytic sample of native-born children in Beirut, Lebanon. The sample was divided into 36 cells, produced by cross classifying in terms of four criteria: sex, private schools vs. public schools, age (5–6 years, 7–8 years, 9–10 years) and religious community (Arab-Christians, Moslems, and

Armenians). With 25 children in each cell, the sample comprised 900 subjects in all. All children of appropriate ages were interviewed in each of several schools, the pupils being subsequently assigned to cells. If more than 25 so obtained *Ss* fell into a given cell, the number was reduced to 25 by discarding the data from *Ss* selected by chance.

Sudanese. This group consisted of 13 girls and 45 boys, all aged 9 and 10, from a private school in Omdurman, the capital of Sudan. Unlike the others, these children gave their answers in writing (in Arabic).

Treatment of Results

For the answers to each question, categories (to be described below) were established to which the majority of replies could be assigned. Each answer was assigned to only one category. Answers that did not fall into any of the defined categories are not represented in the tables of results.

RESULTS

The percentage of each group giving each type of answer to each question is indicated in Table 1. The frequencies for the Lebanese subgroups are shown in Table 2.

The major group contrasts in regard to each question will be discussed in the order in which the questions are listed in the tables. In general, only group differences high in statistical significance are noted. In statements assessing the significance of differences, A-L refers to an American-Lebanese comparison, L-S refers to a Lebanese-Sudanese comparison and A-S to an American-Sudanese comparison.

The Mouth

According to the subjects, the primary uses of the mouth are eating and talking. These two answers make up 100%, 96%, and 92% of the first responses to question 1 for the three groups respectively.

"Talking" as a response to this question is given more frequently by the Americans than by either of the other groups. Table 2 shows that none of the Lebanese subgroups equals the Americans in this regard. Conversely, eating is a response given more often by all segments of the Lebanese sample than by the Americans.

The Sudanese are the only *Ss* to mention drinking as the *first* response to the question "What is the mouth for?" Not shown by Table 1, because it deals only with first-mentioned uses, is the fact that among the Sudanese children a very common response to the question, "What is the mouth for?", is "eating and drinking." This joint mention of eating and drinking occurred among 50% of the Sudanese, but among only 3% of the Americans (S-A $p < .001$) and none of the Lebanese. The frequent mention of drinking in Sudan is probably related to the high temperatures prevailing there, which require a high water consumption.

It is noteworthy that many possible uses of the mouth were not mentioned by any child in any of the three groups. Many possible replies were absent not only as first responses but failed to appear among subsequent responses as well. Among such absent responses were singing, chewing, spitting, shouting, biting, pointing, smiling, laughing, crying, coughing, sucking, blowing, whistling, praying, and cursing. It seems likely that in societies other than the three which were investigated one or more of these would be common word associations with the mouth.

Hands

For all three groups of children, the uses of the hands are more diversified than the uses of the mouth. The most common responses to the question, "What are hands for?" are to hold things, to grasp, to work, to carry, to touch with, to eat with, and to write with. Only the frequencies of eating, writing, working, and playing are examined below.

Eating as something for which the hands are used is mentioned much more often by the Lebanese and Sudanese (by 23% and 33% of the respective groups) than by the Americans (6%) (p of A-S and A-L $< .001$). These figures no doubt reflect differences in eating customs among the groups, holding food with the hands being common among some groups of the Lebanese and Sudanese. Among the Lebanese reference to use of the hands in eating is made more often by the government school pupils who in general belong to a lower economic group than by the private school pupils (Table 2).

In regard to writing and working, the response frequencies of the three major groups are almost identical. Play is mentioned most frequently by the Americans, seldom by the Lebanese, not at all by the Sudanese. Other responses are of such a miscellaneous nature as to be difficult to treat in a comparative manner.

Mother

In all groups, providing food and care for children comprise the major functions of a mother in the eyes of the child. Providing food is most often mentioned among the Lebanese. This response is more common among the younger Lebanese children than among the older, but even among the Lebanese 9–10-year-olds food is mentioned more frequently than by the Americans as a whole. The Sudanese children are unique among the groups in specifically mentioning nursing, 19% of their first-mentioned uses of the mother being to nurse a baby. The complete absence of such responses among the American and Lebanese *Ss* probably reflects a taboo against making references to the breasts in these groups.

Although the general response, "she takes care of us," was common, in all groups there is an al-

TABLE 1
MAJOR GROUP COMPARISONS

Stimulus Words and Use—Categories	Percentage Frequencies of Responses		
	American	Lebanese	Sudanese
Mouth			
eating	61	74	71
talking	39	22	14
drinking	0	0	7
Hands			
eating	6	23	33
writing	11	10	16
working	12	12	12
playing	4	1	0
Mother			
providing food	16	51	7
providing care	62	44	41
nursing	0	0	19
Father			
working, earning money	55	76	10
buying food	1	16	7
providing care	12	6	16
assisting family	14	3	29
Boy			
going to school	16	37	7
working, helping parents	23	24	17
future reference	11	20	12
playing	34	9	0
Girl			
going to school	12	33	9
working, helping parents	28	45	34
future reference	12	11	0
playing	27	5	0
Trees			
food	34	69	52
wood	8	8	10
shade	10	2	12
climbing	12	1	0
Wood			
fire	43	33	5
building	34	17	14
furniture	12	33	21
gates	0	0	17
Dogs			
guarding, watching	16	66	62
barking, biting	9	19	0
hunting	4	8	7
pets, playing	49	4	3
Cats			
catching mice	16	63	53
meowing	6	13	2
pets, playing	52	6	7
Birds			
eating	7	37	16
flying, singing	34	36	36
enjoying	33	8	2
Stones			
throwing, hitting	19	13	2
building	38	73	53
Sun			
drying	0	9	2
warming	25	28	9
light	53	47	67
Sand			
playing	41	18	2
building	21	58	26
Rain			
growing plants	75	46	41
providing water	19	16	12
Gold			
decorative	16	53	67
economic	61	26	2

most complete absence of reference to *specific* motherly acts other than feeding. No child mentioned sweeping, washing dishes, laundering, ironing, sewing, dressing, washing faces, bathing, shopping, putting to bed, arousing, getting off to school, or the superintending of these activities, although some of these would probably have been mentioned if as many responses as was possible had been obtained to each question. Also unmentioned are helping with lessons, moral instruction, reward and punishment, dispensation of justice within the family, love and affection, story-telling, and parent-child play.

Father

If the child's point of view is accepted, the life of a Lebanese father must be a dreary one. He works and earns money in order to support the family. This is the general tenor of all answers. As in the case of the mother, there is no mention of his playing with the child, rewarding or punishing, or of expressing love or hate. So far as the manifest content of the responses are concerned, the Lebanese child seems to have a father who provides for him but who otherwise has little relationship to him. The answers of the American children are similar, but there is a much larger miscellaneous category. In contrast to the Lebanese and the Americans, the Sudanese child mentions the working and earning functions of his father very seldom. Only 10% of the Sudanese answers fall into this category as compared to 55% and 76% of the Americans and Lebanese respectively. Assisting the family and taking care of the family are common answers among the Sudanese but are infrequent in the other two groups.

Boy and Girl

"What is a boy for?" "What is a girl for?" Nearly every answer to these questions can be placed in one of four categories: The function of a

TABLE 2
COMPARISON OF LEBANESE SUBGROUPS

Stimulus	Percentage Frequencies of Responses									
Words and Use—Categories	Moslems	Christians	Armenians	Boys	Girls	Private Schools	Public Schools	5-6 yrs.	7-8 yrs.	9-10 yrs.
Mouth										
eating	81	80	70	79	75	73	75	79	82	70
talking	18	19	29	20	24	27	17	19	17	31
Hands										
eating	28	24	19	21	25	15	30	41	20	10
writing	10	14	10	9	13	15	5	4	13	16
working	22	18	13	11	14	12	12	7	16	27
playing	1	1	4	1	0	1	1	3	2	1
Mother										
providing food	55	42	57	55	47	44	59	65	55	34
providing care	41	53	39	39	49	54	34	28	38	64
Father										
working, earning money	64	77	88	74	78	77	76	70	81	82
buying food	28	11	9	17	15	15	18	23	13	13
providing care	1	4	1	3	3	3	3	2	2	2
assisting family	2	1	0	1	2	2	1	1	1	1
Boy										
going to school	40	37	32	43	30	41	34	40	38	31
working, helping parents	17	18	38	33	26	22	27	16	30	27
future reference	15	28	14	17	23	20	20	8	19	33
playing	8	9	9	9	8	9	9	18	5	3
Girl										
going to school	30	36	32	32	33	35	30	40	33	25
working, helping parents	43	40	49	47	44	41	50	28	44	59
future reference	10	12	10	11	11	12	9	6	13	12
playing	4	5	6	6	5	6	5	11	3	1
Trees										
food	65	70	72	68	70	72	66	63	74	71
wood	10	10	5	8	8	7	9	8	7	12
shade	2	3	2	2	2	2	2	2	2	2
climbing	3	1	1	1	2	1	2	3	1	0
Wood										
fire	40	27	34	33	34	39	28	38	38	24
building	21	22	7	16	17	17	16	21	16	13
furniture	33	48	18	34	32	41	25	16	36	45
Dogs										
guarding, watching	59	68	65	65	63	70	58	37	68	84
barking, biting	17	14	25	16	21	13	24	40	11	5
hunting	11	10	3	9	6	10	6	5	9	8
pets, playing	7	3	1	5	3	3	5	6	4	1
Cats										
catching mice	62	67	54	66	60	71	55	33	69	82
meowing	13	13	13	10	16	9	17	25	10	4
pets, playing	10	7		6	5	4	7	9	4	4
Birds										
eating	40	47	24	36	38	34	40	28	40	45
flying, singing	28	28	50	34	38	35	36	47	35	25
enjoying	11	11	3	6	10	11	5	5	6	13
Stones										
throwing, hitting	18	14	6	13	13	11	14	25	8	5
building	66	71	72	72	67	77	63	50	71	88
roads	4	2	3	3	4	1	5	5	3	2
Sun										
drying	10	11	5	9	9	6	12	10	11	5
warming	32	26	30	23	38	27	30	30	30	29
light	38	46	57	52	42	52	42	34	50	57

TABLE 2—*Continued*

Stimulus	Percentage Frequencies of Responses									
Words and Use—Categories	Moslems	Christians	Armenians	Boys	Girls	Private Schools	Public Schools	5–6 yrs.	7–8 yrs.	9–10 yrs.
Sand										
playing	22	20	13	16	21	28	8	30	18	7
building	50	63	60	66	50	51	64	44	58	72
Rain										
growing plants	45	50	43	51	42	63	30	25	50	64
providing water	17	15	17	15	17	13	18	13	19	16
Gold										
decorative	60	54	52	39	72	49	62	49	52	65
economic	25	28	25	46	12	35	16	17	31	29

boy (or a girl) is (a) to go to school, (b) or to work, help at home, assist mother and father, do chores and run errands, (c) or to grow up in order to do something in the future (become a man or woman, earn money when he grows up, etc., (d) or to play.

Among the Lebanese, going to school comprises 37% of the answers referring to the boy and 33% of the answers referring to the girl. Among the American and Sudanese groups school is mentioned much less frequently, although all of the *S*s were attending school when questioned. Evidence from other studies, to be reported separately, indicates that schooling has a very high value among Lebanese parents and children.

Working (including the helping of parents) stands relatively high in frequency in all groups. In all groups working is mentioned more often as a function of the girl than as a function of the boy. Preparing for future has frequencies roughly equal in all groups, except that among the Sudanese *S*s, most of whom were boys, preparation for the future on the part of girls received no mention.

Play seems to be the outstanding function of American children. It is said to be the purpose of a boy in 34% of the instances and of the girl in 27% of the cases. The corresponding Lebanese frequencies are 9% and 5% (p of A-L for boys and girls each $< .001$). This finding, too, is congruent with other data to be presented in subsequent reports.

No mention of play occurred in the Sudanese answers to these questions. No supporting data are available, but the rarity of Sudanese references to play strongly suggests either that play is relatively infrequent, or that it is disapproved of, or both, among this group.

Trees

Sixty-nine per cent of the Lebanese *S*s state that trees are to provide food. The high proportion of such answers among the Lebanese is no doubt related to the fact that Lebanon has a great variety of food-bearing trees, including olives, almonds, hazel nuts, walnuts, dates, oranges, lemons, grapefruit, bananas, cherries, peaches, plums, and apples. Even pine trees are prized for edible nuts. But American children, living in the same environment, mention the food-producing function of trees only half as often (34%) (A-L $p < .001$). This finding demonstrates again that answers are derived from culture rather than from the physical environment.

The shade-producing function of trees is mentioned more often by the Americans and the Sudanese (10% and 12%) than by the Lebanese (2%) (p of S-L = .01, p of A-L $< .001$). Climbing trees, which is a play function, has a frequency of 12% among the Americans, 1% among the Lebanese, and zero among the Sudanese. Again, since American and Lebanese *S*s were all living in Beirut, it can be assumed that the differences were not produced by differences in natural environment.

Wood

Wood as a fuel is frequently mentioned by the Americans and the Lebanese, but only seldom by the children of Sudan where heated buildings are unknown. Wood as a material used in building is much more familiar to Americans than to the other two groups. No Lebanese or Sudanese homes are built of wood, this scarce material being reserved chiefly for doors and window frames. Sudanese homes are frequently surrounded by an adobe wall, entrance to which is by a wooden gate. It will be noted that wood used for gates is referred to only by the children of the Sudan.

Cats and Dogs

In the Near East and in certain other parts of the world cats and dogs are not pets as they are in Europe and America. In the Near East they are treated in much the same way as are other domestic animals whereas in the Western world they have a special status. The Lebanese and Sudanese

answers reflect the fact that if a dog is kept it is for the purpose of guarding the house and warning against intruders. A dog is occasionally used for hunting. In consonance with these adult notions, guarding and hunting make up 62–66% of the Lebanese and Sudanese responses. Among the Lebanese, to play with a dog and to pet a dog are referred to only by the younger children, among whom such answers occur in 4% of the cases. Attitudes towards dogs do not differ as between the Lebanese Moslem and Christian groups (see Table 2).

In contrast, 49% of the American answers indicate that dogs are pets and are to be played with (A-L $p < .001$). Among the Americans such answers are as frequent among the older children as among the younger.

The differences among the groups in regard to cats is similar to that prevailing in reference to dogs. For the Lebanese and Sudanese the primary feline purpose is the catching of mice and rats. Occasionally the catching of insects by cats is mentioned. When specific insects are mentioned, they are usually cockroaches. Reference to cats as animals to play with or to pet occurs infrequently in these groups (6–7%) but 52% of the American answers fall into this class ($p < .001$).

Birds

For some children in all groups, birds are to fly, to sing, and to be put in a cage. But 37% of the Lebanese *Ss* state that birds are to be killed and eaten. This reflects the Lebanese custom of shooting and eating many birds which Americans do not classify as game birds. The reader may note again the prevalence of responses referring to food on the part of the Lebanese. The Sudanese mention birds as food less frequently (16%) than the Lebanese, but more frequently than the Americans (7%) (A-L $p < .001$).

Stones

In all groups stones are primarily to build with, 38%, 73% and 53% of the answers, respectively, being in this category. This answer is not surprising in countries in which stone has been the most widely used building material for thousands of years.

"Stone" was included in the stimulus list in part because stones can be used aggressively. In this regard only small differences appear among the groups. Using stones for throwing or hitting makes up only 13% of the Lebanese, 19% of the American, and 2% of the Sudanese responses. That the frequency of this answer is low among the Sudanese seems to be due to the greater age of the Sudanese subjects. Age is a major variable affecting aggressive verbal responses in regard to stones (see Table 2). Among the Lebanese responses, throwing and hitting make up 25% of the responses of the 5–6 year olds to the word "stone." Beyond this age such responses become infrequent in both Lebanese and American groups, apparently indicating strong disapproval of overt aggression.

The Sun

As might be expected the sun's predominant functions are to provide warmth and light. On this item one of the major group contrasts lies in the fact that the Sudanese *Ss* referred to the heat-providing function of the sun much less often than the other groups (9% as contrasted to 25–28%) (A-S $p < .001$). Apparently, the Sudan is ordinarily so hot that the child thinks the sun's contribution is not useful!

Another service provided by the sun is that of drying clothes. This is mentioned by 9% of the Lebanese, 2% of the Sudanese, and none of the Americans. Presumably, more clothes are dried at home in the former groups. Table 2 indicates that among the Lebanese this answer is more prominent among the very young (10%) than among the oldest group (5%) and more frequent among the poorer than among the more well-to-do (12% and 6%).

Sand

Because of very extensive building operations which are under way in Beirut, construction work is familiar to all Beirut children. The use of sand in building is mentioned by 58% of the Lebanese. American children living in Beirut mention construction much less often (21%) ($p < .001$).

One of the major group differences in regard to sand involves play. Playing in sand is mentioned by 41% of the Americans, 18% of the Lebanese (A-L $p < .001$), and 2% of the Sudanese. The generally high rating of the Americans on play responses has been noted in connection with several earlier questions. The low play rating of the Sudanese *Ss* on this particular question is in part due to their ages, since sand play is uncommon at ages 9 and 10. Among the Lebanese, play as a use for sand has a frequency of 30% at ages 5–6 and of only 7% at ages 9–10. It will be noted that the frequency for American children for ages 5–10 combined is higher than the frequency at ages 5–6 among the Lebanese.

Rain

Rain provided no group differences of special interest. In all groups, the most common use attributed to rain is that of aiding the growth of plants. This response is highest among the Americans. The other two groups gave many more answers which fell into the miscellaneous or unclassified category.

Gold

Almost all answers concerning gold fall into two classes: Those referring to decorative uses such as the making of rings, bracelets and earrings, and those referring to economic uses, such as statements to the effect that gold is to sell, or to buy things, or to make money, or to make one rich.

The Sudanese frequently mention decorative uses and seldom mention economic ones. Americans give a reverse picture, economic responses predominating. The Lebanese have intermediate response frequencies. These differences are probably related to variations in the use of gold ornaments among the three groups, and to variations in acquaintance with the monetary value of gold.

In Lebanon, chocolate candies are frequently wrapped in gold foil. As a result, among the younger Lebanese children a not infrequent answer to "What is gold for?" is "To wrap candy in."

Comparisons of Lebanese Subgroups

In the foregoing comparisons among the three main groups, reference has been made to Lebanese subgroups only when it clarified the main group contrasts. Let us now examine in more detail the data contained in Table 2.

While Lebanon is noted for its cultural diversity it appears from Table 2 that the children of Moslems, Armenians, and Arab Christians have much in common. The first three columns of Table 2 present data for these groups. In most instances, the percentages for Moslems and Arab Christians, who have lived side by side for centuries, are almost identical. Since the majority of Armenians in Lebanon have immigrated to this country since the Turkish massacres of 1914, they have had a relatively short time in which to become acculturated. Their responses differ somewhat from those of the other Lebanese groups. But they diverge much less from the Moslems and Arab Christians than do the Americans and the Sudanese.

So far as the answers given by the two sexes are concerned, again there is a general similarity at the age levels represented in this study. Only a few significant sex differences appear. In regard to the sun, girls stress warmth more than do boys (38% vs. 23% $p < .001$). Playing in sand is more often mentioned by the girls (21% vs. 16% $p < .001$). The largest difference appears in regard to gold. Decorative uses are given by 72% of the girls and only 39% of the boys ($p < .001$).

Age affects the frequency of responses to nearly every question. Thus, eating as a function of the hands decreases with age, and writing increases. Play in general decreases, whereas work increases. The reader with a special interest in age differences will wish to examine Table 2 in some detail. It will not be feasible to comment on age changes in respect to each item. Concepts of use strongly reflect developmental age. This fact underlines the importance of equating for age any groups which are to be compared. The reader should bear in mind that the American and the Lebanese groups were equated for age composition. The fact that the Sudanese group consisted entirely of 9- and 10-year olds has been taken into account in making all comparisons.

Hypotheses Concerning the Lebanese Child

The data have been seen to reveal a considerable number of differences among the three major groups of *S*s. It remains to attempt to characterize in a general way the answers of each group. Such a general characterization is warranted only for the Lebanese and the American groups in view of the small size of the Sudanese sample and the fact that it represented only one age level.

It has already been observed that the Lebanese answers seem to show a considerable concern with food. Common answers include the following: mother feeds us, father provides food, the mouth and the hands are used for eating, trees produce food, and birds are to be eaten. The answers taken by themselves do not show why food is so important. There is no reason to believe that the *S*s suffered from malnutrition. Actually, the majority of Lebanese children appear to be very well fed. Observations other than those presented here support the tentative suggestion that food preparation and eating have for the Lebanese a pleasure function that is comparable to the positive valuation given by some societies to other activities such as music, conversation, or visual esthetic experiences.

In the second place, the Lebanese responses seem to indicate that industriousness is highly valued. Note the following common answers: father and mother work, girls and boys work, the hands are for work, dogs work by guarding the house, cats work by catching mice and rats. Trees and birds provide food, which may be thought of as a work function. Stones and sand supply building materials. Much of the world seems to be oriented toward utilitarian ends.

Concern for language facility and scholastic performances seems evident, particularly among the private school pupils. The mouth is for talking, the hands for writing, and one of

the chief functions of both boys and girls is to go to school.

Few other emphases are observable in the Lebanese answers. There is some mention of play in connection with questions about boys and girls, and the question about sand, but little or no mention of it in reference to parents, hands and mouth, or animals. There is no mention of art, of music, of athletics, or of sensory pleasures other than eating. There is little indication of valuation of emotional life or of group cohesion. There are no references to competitiveness in areas other than school achievement. There is nothing to suggest that religious observances or religious affiliations are highly significant.

Hypotheses Concerning the American Child

The present data, it is important to remember, are restricted to American children in Lebanon, and more specifically, to those attending a specific progressive school. The parents are nearly all college graduates, and most seem to have accepted the doctrine of permissiveness. While many groups of children living in America probably would answer as do the American group in Lebanon, not all groups of children in America would do so.

The Beirut colony of American children (at ages 5–10) may be characterized, from their answers, as being greatly concerned with play. To a much greater extent than among the Lebanese *Ss*, the hands are for play, boys and girls have play as their purpose, trees are for shade and climbing, dogs and cats serve as pets and playmates, birds are to be enjoyed, and sand serves a play function. In contrast to the seriousness of the Lebanese children, life for the American children in Beirut would seem to constitute a perpetual vacation.

Positive mention of play is associated with low frequencies of reference to work and to duties. Going to school, and helping parents are reported as functions of boys and girls much less often by Americans than by the Lebanese. The function of the father as a bread-winner is less frequently cited by the American group. As noted previously, dogs and cats are for fun and are seldom thought of as utilitarian animals. Similarly, the use of stones and sand as building materials is stressed much less by the American group. Apparently the American child sees the world more in terms of pleasure and less in terms of duty and work than does the Lebanese child.

This difference appears not to be due to any appreciable extent to differences in social class. If the American group is compared with the Lebanese private school group (Table 2) rather than to the entire Lebanese sample the differences remain substantially unchanged. It is also true,that Americans diverge as much from Lebanese Christians as from Lebanese Moslems, these two groups scoring very similar percentages in the response categories to every question (see Table 2).

The categories listed in Table 1 include only the most common answers. The responses that are not represented in this Table were given by only a few children. They may therefore reasonably be called individual or unconventional answers, and may represent individuality or originality. If one adds for each question the percentages of Americans and Lebanese giving answers in the categories of Table 1, one finds that to fourteen of the sixteen questions the Lebanese gave a larger proportion of common, or stereotyped answers. The American answers, in general, show more variety. This finding may represent the result of permissiveness, it may point to more varied experiences, or it may be the result of direct encouragement of individual interests and modes of thought. This difference seems to the investigator to be among the more interesting ones revealed by the study.

SUMMARY

A technique for the study of children's concepts of the uses of common objects has been presented which is believed to be universally applicable. Its use has been illustrated with American and Lebanese groups and to a lesser extent with Sudanese *Ss*. The study has demonstrated that in regard to universally present objects, such as hands and mouth, dogs and cats, sun, sand and stone, children of different cultures differ in their ideas as early as five years of age. It has also demonstrated that the test responses enable one to formulate hypotheses concerning the interests, values, and activities which are being developed in different groups of children. The method is believed to be a useful addition to the research techniques available for the study of cultural differences in child development.

Received July 30, 1956.

AUTHORITARIANISM, INTOLERANCE OF AMBIGUITY, AND RIGIDITY UNDER EGO- AND TASK-INVOLVING CONDITIONS[1]

THEODORE MILLON

Lehigh University

IN SPITE of extensive research dealing with the connection between measured authoritarian attitudes and behaviors referred to in the literature as rigidity and intolerance of ambiguity, the existence of a true covariance is still in doubt and the antecedent conditions necessary for such a covariance remain unclear. Early research by Rokeach (9) and Frenkel-Brunswik (1, 7) obtained positive relationships between California E- and F-scale scores and intellectual and perceptual measures of rigidity and intolerance of ambiguity. Numerous subsequent studies have failed to substantiate their findings. The view that an atmosphere of ego-involvement or personal stress may be a necessary condition for obtaining such relationships has been recently advanced (3). Unfortunately, experiments designed to test this view have merely contributed additional conflicting evidence (5, 6). The problem has been complicated further by suggestions that induced ego-stress in itself is sufficient to produce rigid and ambiguity-intolerant behavior in experimental subjects (*Ss*), independently of supposedly predisposing personality traits. Research regarding this point has likewise been equivocal (3, 4, 6, 10).

The number and complexity of contradictory experiments in this area make it impossible to distinguish common factors differentiating studies that obtain a positive relationship between the California scales and rigidity and ambiguity-intolerance from those in which such a relationship has not been found. Some of the difficulty can be attributed to differences in experimental procedures used to induce ego-stress and perhaps even more from a failure to evaluate methods for operationally defining rigidity and intolerance of ambiguity. Ego-involvement has usually been administered to large groups with no attempt made to evaluate objectively the effectiveness of experimental instructions. Tasks selected to measure rigidity and intolerance of ambiguity have been many, most having little more in common than the name. Few of the methods have experimentally differentiated the behavioral effects of ambiguity-intolerance from those of rigidity, and the majority appear not to measure variables relevant to the California theory of authoritarianism (1).

The present study was designed to introduce some clarity by: (*a*) administering ego and task orientations individually, and running postexperimental checks on their effectiveness, and (*b*) operationally defining and distinguishing between intolerance of ambiguity and rigidity in a manner appropriate to theoretically posited authoritarian characteristics.

In terms of the experimental tasks constructed for the present research, intolerance of ambiguity refers to the tendency to respond to previously unexperienced and unstructured stimuli with a subjectively structured response set. Rigidity refers to the tendency to resist modifying an established response set when the stimulus conditions under which it was established have been changed. To assure that resistance to change is attributable to personality factors and not to initial differences in the habit strength of sets, experimental control of set strength among *Ss* is necessary. This control is also essential in order to differentiate the behavioral effects of ambiguity-intolerance and rigidity. For example, if *Ss* in an unstructured stimulus situation are given an *equal* number of trials to establish a response set, *Ss* high in ambiguity-intolerance may be expected to establish a stronger set than *Ss* low in ambiguity-intolerance because of their tendency to acquire it earlier and apply it more consistently. Under these conditions, it is psychologically meaningful to speak of rigidity, as gauged by resistance to extinction, only when differences in set strength among *Ss*

[1] This paper is adapted from a dissertation submitted in candidacy for the Ph.D. degree at the University of Connecticut. The writer is indebted to Maurice L. Farber, thesis advisor, and W. L. Jenkins for assistance in the present formulation. A report of this study was given at the APA meeting, 1954.

are equalized. If rigidity is to serve as a motivational concept, it must contribute to resistance to extinction above and beyond simple habit strength. Unfortunately, many of the operational measures of rigidity currently in vogue fail to assess and control for differences in initial set strength.

In light of these considerations, the following questions were posed for experimental analysis: 1. Do authoritarians subjectively establish a set or norm in a previously unexperienced and unstructured situation more rapidly than equalitarians? 2. *After establishing an equally strong response set*, are authoritarians more resistant to modifying their set than equalitarians when stimulus conditions are changed? The experiment was further designed to answer these questions: 3. Does induced ego-involvement increase relationships among measures of authoritarianism, ambiguity-intolerance, and rigidity? 4. Does ego-involvement in itself produce greater rigidity and intolerance of ambiguity than nonego-involvement?

METHOD

The *S*s whose scores on the standard 30-item Form 40-45 California F scale (1) were known, were asked to participate in a one-session laboratory experiment involving two series of judgments of autokinetic movement. The autokinetic phenomenon satisfied the requirements of unstructuredness and freedom from related prior experiences. The fact that positive results emerged from the only previous study relating the California scales to autokinetic judgments (2) gave further evidence of its appropriateness as a research instrument.

Series 1 consisted of judgments of the direction and distance of movement of a one-light autokinetic stimulus and was designed to measure the speed of norm or set formation; scores were represented by the number of stimulus presentations necessary before *S* attained a standard experimental criterion of establishing a response set. The criterion required that 10 out of 13 consecutive judgments be reported as moving in only one of four primary directions (up, down, left, right) with a magnitude varying not more than one inch from one another. On the basis of previous research, the assumption is warranted that barring significant changes in the stimulus field, *S*s would retain this response set as a frame of reference for subsequent judgments. Series 1 was ended for each *S* whenever a sequence of his judgments met the criterion. Although the number of trials to reach the criterion varied among *S*s, the same degree of consistency in judgments was required of all *S*s as the basis for terminating Series 1. Set strength among *S*s was thus operationally equated at the conclusion of Series 1, and since the second series followed immediately, *S*s began Series 2 with equal habit strengths in their set for the one-light situation. During Series 2, a second light was added to the stimulus field in conjunction with the first light. The *S*s were simply told at this point to continue judging the distance moved by the first light. Preliminary experiments (8) had found that the simultaneous viewing of two lights typically curtailed movement from that perceived under one-light conditions. Series 2 consisted of 50 judgments and was designed to determine the extent to which *S* shifted from his established one-light norm or set when viewing the two-light stimulus.

TABLE 1
F-SCALE GROUPS SELECTED FOR EXPERIMENTATION

Group	*N*	Approximate F-Scale Centiles	Range of F-Scale Scores	Group F-Scale Mean
L_{10}	14	0–10	1.6–2.9	2.52
L_{20}	15	11–20	3.0–3.1	3.05
M	13	45–55	3.5–3.8	3.67
H_{20}	14	80–89	4.3–4.5	4.35
H_{10}	13	90–99	4.6–5.2	4.84

Two operational measures were obtained in the autokinetic sequence: intolerance of ambiguity by the speed the norm was formed, and rigidity by the relative distance toward zero *S* shifted from his one-light norm during the two-light series.

Approximately half of the *S*s received instructions designed to create an ego-involving atmosphere, i.e., the impression that the tasks reflected their abilities and intelligence; the other half received instructions to create a task-involving orientation, i.e., the impression that the experiment was designed to try out apparatus for subsequent research on vision. A posttest questionnaire was administered to check on the effectiveness of the induced orientations at the close of the laboratory session. Detailed descriptions of the procedure and questionnaire may be found elsewhere (8).

Subjects

The *S*s were 69 students enrolled in an introductory psychology course. They were assigned at random to one of the two experimental orientations from five divisions of a California F-scale distribution of 162 students. Table 1 presents the centile and F-scale score range for these divisions. Nine *S*s failed to establish a norm within the experimentally predetermined maximum of 150 Series 1 stimulus presentations. Norm formation scores for these *S*s were arbitrarily set at the maximum of 150. Of the remaining 60 *S*s who had met the criterion of norm formation prior to the 150th trial, two subsequently complained of illness and asked to be excused during Series 2.

RESULTS

An answer to the question of whether authoritarians are more ambiguity-intolerant and rigid than nonauthoritarians can be gathered from Tables 2 and 3. The analysis of scores on trials to form a norm, contained in Table

TABLE 2
ANALYSIS OF VARIANCE OF TRIALS TO FORM A NORM
(Scores transformed to reciprocals)

Source	*df*	Mean Square	*F*	*p*
Orientation	1	2.00	.61	n.s.
Authoritarianism	4	33.95	10.77	.001
Interaction	4	1.05	.32	n.s.
Within	59	3.30		

TABLE 3
ANALYSIS OF VARIANCE OF PERCENTAGE OF SHIFT FROM NORM

Source	*df*	Mean Square	*F*	*p*
Orientation	1	13,922.5	26.05	.001
Authoritarianism	4	1,293.2	2.42	.06
Interaction	4	657.9	1.23	n.s.
Within	48	534.3		

2, indicates a clear-cut relationship ($F = 10.77$, $p < .001$) between intolerance of ambiguity and authoritarianism. The relationship between authoritarianism and percentage shift from norm, noted in Table 3, yielded an F of 2.42 ($p < .06$). Although suggestive, this finding does not offer clear support for a correspondence between the California scale and rigidity.

The second major question of the study was whether induced ego-involvement would increase relationships among the measures. Figure 1 illustrates the mean number of trials to norm formation for each of the five F-scale groups under both ego and task orientations. The relationship between rigidity and F-scale scores in terms of percentage of shift is portrayed in Figure 2. It can be seen from this figure that the relationship between F-scale scores and rigidity is more marked in the ego-oriented group. Task-oriented authoritarians, for example, evidence little more rigidity than task-oriented equalitarians; the difference in the ego-oriented condition, in contrast, is quite pronounced. The interaction *F*s of .32 and 1.23 noted in Tables 2 and 3 indicate, however, that orientation factors have not been demonstrated to play a statistically significant role in obtaining the generally high relationships found between authoritarianism, ambiguity-intolerance, and rigidity.

The effect of induced ego-involvement in producing increased ambiguity-intolerance and rigidity among experimental *S*s is also illustrated in Figures 1 and 2. Although the 9.8 mean difference between ego- and task-oriented groups in trials to norm is suggestive, it is clear from Table 1 that the difference ($F = .61$) does not attain statistical significance. In contrast, a 21.1 per cent greater resistance to shift on the part of the ego-oriented group is significantly greater ($F = 26.05$) than that shown by the task-oriented *S*s. This difference is independent of F-scale scores. Figure 2 clearly shows that all of the ego-oriented F-scale groups evidenced considerably more rigidity than comparable groups receiving the task orientation.

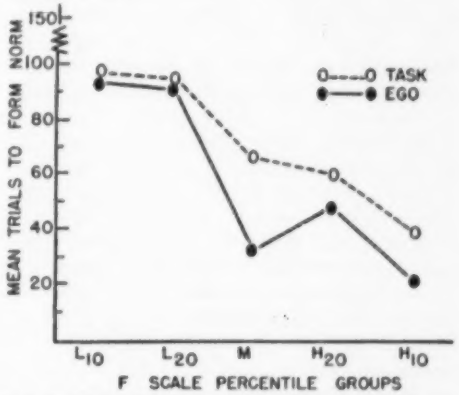

FIG. 1. MEAN NUMBER OF TRIALS TO FORM NORM UNDER EGO- AND TASK-ORIENTING CONDITIONS

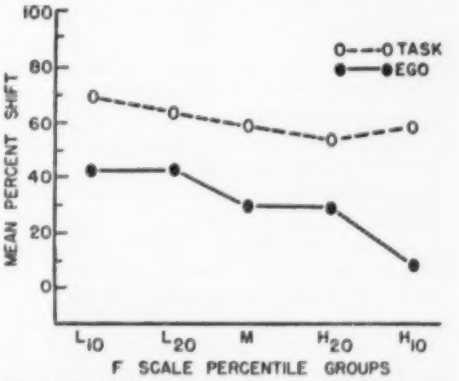

FIG. 2. MEAN PERCENT OF SHIFT TOWARD ZERO FROM ONE-LIGHT NORM UNDER EGO- AND TASK-ORIENTING CONDITIONS

The introspective self-involvement ratings of the posttest questionnaire established a significant difference in the direction expected ($F = 11.56$) between ego- and task-oriented groups. The questionnaire indicated further, that task-oriented *S*s showed an appreciable degree of concern regarding their personal performances in spite of the experimental instructions used. No systematic difference was found among the F-scale groups in receptivity to the two testing orientations.

DISCUSSION

The fact that authoritarians formed norms in the unstructured autokinetic situation more rapidly than nonauthoritarians under both ego- and task-orienting conditions clearly supports Frenkel-Brunswik's position (1, 7) concerning the characteristic ambiguity-intolerance of authoritarians. The general finding that authoritarians modified their established norms to an appreciably lesser degree than nonauthoritarians, offers additional support for the association between rigidity and authoritarianism asserted by the California group. There appears, however, to be a limitation to this covariance. Although authoritarians manifested greater rigidity than nonauthoritarians in the ego-orienting atmosphere, no such difference was evidenced in the task group. Though perhaps a favored mechanism it would appear that rigidity in the authoritarian is partially contingent upon variables associated with ego-involvement. What factors mediate this relationship is problematic.

Although induced ego-involvement cannot be considered to have played a statistically significant part in contributing to relationships found between authoritarianism, ambiguity-intolerance, and rigidity, it is of interest to note that had the experiment been conducted under task-orienting conditions only, the relationship between rigidity and the F scale would have been reported as not significant ($F = .462$). In contrast, had the experiment been run under ego-orienting conditions, the results would have been considered significant ($F = 3.40$; $p < .03$). Although minimal faith should be put in an hypothesis when the statistic called for by the experimental design turns up a nonsignificant effect, it seems wise, in light of the above facts, to reconsider the possible contribution of ego-involvement in the rigidity-authoritarian relationship. This caution appears particularly justified since the appreciable degree of ego-involvement found among our task-oriented *S*s made it impossible to evaluate experimentally the full effects of the ego-involvement variable. Had the difference between the two orientations been more marked, its contribution to authoritarian rigidity may have been more clearly established. In any event, further study of ego-involvement in this context would appear warranted.

A major question underlying the present study was whether personal stress or ego-involvement would increase the general level of ambiguity-intolerant and rigid behavior among *S*s regardless of their standing on the F scale. Although the influence is a dubious one as regards intolerance of ambiguity, there can be no doubt as to its effects on behavioral rigidity. Whether ego-involvement was in itself sufficient to bring about the increase cannot be readily determined. The following speculations regarding the connection may nevertheless be offered: *S*s who were ego-involved with respect to their one-light norm were motivated to ward off stimuli that conflicted with or questioned the adequacy of this norm. Because of its tendency to curtail the impression of movement, the second light had the effect of challenging the *S*'s established one-light norm. Inasmuch as past judgments presumably reflected intelligence and ability, *S*s, in an attempt to safeguard self-esteem and reassure themselves of the adequacy of their one-light performance, warded off the visual influence of the second light. Task-oriented *S*s, not feeling a personal identification with their performances in the one-light situation had little or no reason to retain or defend their norm. The thesis suggested then is that ego-involvement in the context of the laboratory tasks of this experiment served to motivate *S*s to ignore or reject what was antithetical to their first-formed impressions. It would therefore seem plausible to hypothesize that the effect of a norm-changing stimulus would vary with its acceptability to an individual and that the more ego-involved an individual is with respect to his norm, the more intense will be the operation of his defenses against stimuli that challenge it.

The above interpretation of rigidity as a defensive reaction to the threat posed by the

two-light stimulus adds an additional interpretive problem. Since authoritarians evidenced considerably greater rigidity than nonauthoritarians in the ego-orienting condition, the issue arises as to whether this difference might be better attributed to greater ego-threat experienced by authoritarians than to greater rigidity manifested by them in response to equivalent degrees of threat. Because of its highly structured and limited nature, the posttest questionnaire offered no effective clues as to the validity of this hypothesis. If authoritarians were in fact more threatened than nonauthoritarians when the second light was presented, their greater resistance to change could be considered commensurate with the degree of threat they experienced. If this speculation is valid, authoritarians would be characterized better by their susceptibility to intense threat when their personal norms are questioned than by their "generalized mental rigidity."

Space does not justify a detailed review of how results of the present study agree or disagree with earlier research. Obviously, there are striking differences between experimental techniques used for measuring intolerance of ambiguity and rigidity. The fact that the present ego- and task-orienting instructions were given person-to-person rather than in a group setting, as is more usual, is a further obstacle to comparison. Until additional unequivocal data are available to reconcile contradictory evidence from previous studies, it seems wisest to view new positive results concerning the relationship of rigidity and tolerance of ambiguity to authoritarianism with caution. Clearly, these behaviors are influenced not only by personality factors, but also by the tasks utilized for measurement and the situational atmosphere within which the *S* must respond to them.

SUMMARY

Individual laboratory sessions involving judgments of autokinetic stimuli were conducted with 69 college students whose standings on the California F scale were known; approximately half received ego-orienting instructions, half received a task orientation. Two behavioral scores were obtained during the experimental sequence: (a) the number of trials to form a norm while estimating direction and distance of a one-light autokinetic stimulus, as a measure of intolerance of ambiguity, and (b) the percentage of appropriate shift from the *S*'s established one-light norm upon the introduction of second light, as a measure of rigidity.

Analyses of variance of these measures produced the following findings:

1. (*a*) An *F* beyond the .01 level of confidence between authoritarianism and intolerance of ambiguity.

(*b*) An *F* attaining the .06 level between authoritarianism and rigidity. The relationship was clearly significant in the ego-oriented group, but not in the task-oriented group.

2. An *F* beyond the .01 level of confidence signifying greater rigidity in the ego- than the task-oriented group.

3. Ego-involvement did not significantly increase relationships among the three experimental measures, nor did it produce a general increase in ambiguity-intolerance.

The results of the study were felt to be consonant with positions taken by Frenkel-Brunswik regarding the typical rigid and ambiguity-intolerant behavior of authoritarians. Limitations to this relationship were noted and alternate hypotheses suggested.

REFERENCES

1. ADORNO, T. W. *et al. The authoritarian personality.* New York: Harper, 1950.
2. BLOCK, J., & BLOCK, J. An investigation of the relationship between intolerance of ambiguity and ethnocentrism. *J. Pers.*, 1951, **19**, 303–312.
3. BROWN, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 469–476.
4. COWEN, E. L. The influence of varying degrees of psychological stress on problem-solving rigidity. *J. abnorm. soc. Psychol.*, 1952, **47**, 512–519.
5. DAVIDS, A. The influence of ego-involvement on relations between authoritarianism and intolerance of ambiguity. *J. consult. Psychol.*, 1956, **20**, 179–184.
6. FRENCH, E. G. Interrelation among some measures of rigidity under stress and nonstress conditions. *J. abnorm. soc. Psychol.*, 1955, **51**, 114–118.
7. FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, **18**, 108–143.
8. MILLON, T. The role of situational variables in perceptual behaviors characteristic of authoritarians. Unpublished doctor's dissertation, Univer. of Connecticut, 1954.
9. ROKEACH, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259–278.
10. SMOCK, C. D. The influence of stress on "intolerance of ambiguity." *J. abnorm. soc. Psychol.*, 1955, **50**, 177–182.

Received July 31, 1956.

COGNITIVE COMPLEXITY AND ASSIMILATIVE PROJECTION IN ATTITUDE CHANGE

RICHARD M. LUNDY AND LEONARD BERKOWITZ

University of Wisconsin

AMONG the factors that have been investigated in studies of attitude change (e.g., 4, 5) are the personality of the person being influenced, and the source of the attitude change material. The present study was designed to test hypotheses in each of these general areas.

According to the first hypothesis, attitude change is more likely to occur in "cognitively complex" persons than in "cognitively simple" persons. The concept of cognitive complexity-simplicity stems from the theoretical orientation of Kelly (6), which assumes that each individual possesses a system of constructs for understanding and predicting his world. In comparison with a simple system, a complex system differentiates more sharply among ideas. It distinguishes among a greater number of different aspects of people or situations, employing more functional constructs and showing more variety in the relationships among them.

A basic assumption in the present study is that the ideas expressed in a communication have meaning for its recipient and can therefore be considered by him only if they are relevant to his existing personal constructs. The complex system, having more functional constructs, furnishes greater opportunity, statistically, for a new idea to be considered. Once considered, the new idea may be integrated into the person's construct system to result in attitude change. In contrast, the cognitively simple system, equipped with fewer functional constructs, allows fewer opportunities for new ideas to come in relation to a personal construct and to be integrated into it. New *words* may be manipulated, but unless the new idea is related to a relevant personal construct it can effect no change in the person's construct system and, in a sense, has no meaning for him.

This formulation resembles the Gestalt hypothesis that attitude changes normally occur because the "material to be judged is seen in a new light and has consequently changed its meaning" (7, p. 19). Both views emphasize the importance of the meaning of the communication to the recipient. The present formulation sets up special conditions for "change in meaning" by stating that the new material must have relevance for an individual's personal constructs before it can be "seen in a new light."

The remaining hypotheses concern assimilative projection or assumed similarity (3, 8) as relating the source of attitude change material to the person being influenced. Assimilative projection refers to the degree to which a person perceives other persons (the communication source in this case) as similar or dissimilar to himself. If a person perceives the source of a new idea as similar to himself, he is likely to see the idea also as similar, in some ways, to his own ideas, and thus consider it "correct." He should therefore be more susceptible to influence by a particular source group if he perceives that source group as similar to himself. In the present study, two source groups are used—authority and peers. Hypothesis II states: Persons who are susceptible to peer influence are more likely to perceive peers as similar to self than persons who (*a*) are negative to peer influence, or (*b*) are susceptible to authority influence. Hypothesis III states: Persons who are susceptible to authority influence are more likely to perceive authority figures as similar to self than persons who (*a*) are negative to authority influence, or (*b*) are susceptible to peer influence.

METHOD

Attitude Change Measure

During another investigation by the authors (1) a 132-item attitude questionnaire was administered to students (male and female) in two intermediate psychology courses. Scattered throughout the questionnaire were items from three modified Thurstone scales: attitudes toward the movies, tariff, and the punishment of criminals. The students were told the questionnaire involved an investigation into the "psychology of normal people" concerned with their attitudes on "every-day matters."

One month after the original testing session, four alternate forms of the second questionnaire were administered to the two classes. These questionnaires

were introduced with the following orally-delivered statement:

> Many investigations into the psychology of attitudes have dealt with issues generally given little thought or attention by people. We want to study attitudes that have been given some thought. In order to stimulate your thinking on these issues we have provided you with the opinions of other people. For each of the three attitude issues, first read the opinions expressed by these other people, then indicate your opinion on each item.

Half of the questionnaires passed out to the classes presented the opinions supposedly expressed by peers with this introduction:

> A national public opinion organization recently conducted a survey of the opinions of college students in the universities of the Big Ten... The survey showed a fairly high degree of consensus among the college students on several issues. On the following pages we have presented the attitudes of the great majority of college students.

The remaining half of the questionnaires contained opinion statements supposedly expressed by generals. The students were told in these questionnaires:

> A writer for a national magazine recently interviewed Generals Bradley, Clark, and Taylor on certain issues of national interest. The generals were in complete agreement on these issues. On the following pages we have presented their attitudes.

In contrast to the questionnaire administered in the first session, the items pertaining to a given issue, i.e., either movies, tariffs, or the punishment of criminals, were presented together on a single page. Before each page containing the given attitude scale was a page describing the opinions of either the peers or the generals on this issue. To illustrate, the items in the attitudes-toward-movies scale were preceded in some of the booklets by a page containing this statement: "The great majority of the college students expressed favorable attitudes toward movies. They felt that movies often were very fine entertainment and frequently contributed to the education of the moviegoer." For each attitude scale, half of the booklets contained favorable opinions by either the peers or generals while the remaining booklets contained unfavorable statements. The same opinion rationale was used for both the peers and generals. After reading the introductory statement to each issue, the subjects (*Ss*) indicated their opinions on the given attitude items by marking an IBM sheet.

Cognitive Complexity-Simplicity

Cognitive complexity-simplicity was measured by the Role Construct Repertory Test (RCRT), an instrument suggested by Kelly and used by Bieri (2) in his study on cognitive complextiy. The test consists of a grid across the top of which *S* writes the names of 22 people personally known to him, who fit in prescribed role categories. The *S* is asked to consider three of these people at a time and to decide in what important way two of them are alike and different from the third. Different groups of three are considered successively by *S* until he has listed along the side of the grid a certain number of personal constructs derived from the original list of people. For each construct thus derived, the *S* then considers all the rest of the people in his list and places a check mark under the names of those people perceived as similar to the original two in terms of the construct. In this way the grid is formed with checks or voids which indicate the particular manner in which the *S* perceives similarities and differences within a group of people on the basis of his personal constructs.

By comparing each construct row of check marks with each other construct row, one can determine their degree of similarity. If two rows have check marks under the same names, the constructs are functionally equivalent. When many of the construct rows for an *S* are alike, he is considered as cognitively simple. If he produces few rows that are identical or similar to any other row, he is regarded as having high cognitive complexity. The scores for cognitive complexity-simplicity were obtained by using the first 10 rows and first 10 columns in the standard RCRT.[1] Each row was matched with each other row. When a matching was identical, a score of three was given; when a matching was identical except for one, a score of two was given; and when except for two, a score of one. A high score indicates cognitive simplicity; a low score, cognitive complexity.

Assimilative Projection

One of the persons which the *S* is instructed to use in the RCRT is himself. After the grid is completed, the column of checks and voids under the *S*'s name indicates how he perceives himself in regard to his own constructs. By comparing this self perception with his perception of other people as indicated by his test responses, his degree of perceived similarity, i.e., assimilative projection, can be ascertained. To find assimilative projection with peers, four figures were selected from the first ten on the grid to be matched with the self: (*a*) a brother, or brother-like person, (*b*) a sister, or sister-like person, (*c*) the present closest opposite-sex friend, and (*d*) the present closest same-sex friend. The assimilative projection score for peers was made up of the number of times a check or void under the self matched a check or void under each of the peers considered separately.[2]

The assimilative score for authority figures was obtained the same way, using: (*a*) the mother figure, (*b*) the father figure, (*c*) the minister, priest, or rabbi with whom personal feelings about religion might be discussed, (*d*) the physician, (*e*) the teacher who influenced *S* most when in the teens, (*f*) the *S*'s employer, supervisor, or officer during a period of great stress. In scaling the items, it was found that mother figure, father figure, and employer figure, individually, correlated very little with the total score. In the subsequent analysis of data, therefore, only the religious person, the physician, and the teacher were used in determining the score on similarity to authority. The latter two were not among the figures included in determining the cognitive complexity score.

---

[1] To facilitate scoring, the first ten rows and columns were used instead of the complete grid.

[2] An analysis of the data revealed a slight negative relationship between assimilative projection on peers and cognitive complexity. Since both these variables are positively related to attitude change, this slight negative relationship indicates that the tests of the two hypotheses are independent of each other.

TABLE 1

COGNITIVE COMPLEXITY-SIMPLICITY IN RELATION TO SOURCE OF INFLUENCE AND TYPE OF ATTITUDE CHANGE

Type of Attitude Change	Mean Cognitive Complexity-Simplicity*	
	Generals (*N*) as source	Peers (*N*) as source
Little change	(9) 16.44	(6) 35.50
Variable change	(5) 16.20	(9) 14.78
Susceptible	(7) 13.43	(9) 14.89
Negative change	(6) 8.83	(12) 12.58

Analysis of Variance

Source of Variation	*DF*	*MS*	*F*	*p*
Source of influence	1	200.16	—	—
Type of attitude change	3	570.56	5.0	.05
Interaction	3	297.13	2.6	—
Within groups	55	114.19		

* A high score represents greater simplicity.

Analysis of Data

Change scores from the first to the second testing were determined for each of the three scales. These raw scores were then converted into standard scores with a mean change score of 20 for each scale, and the three scale scores added to obtain a total standard susceptibility score. Since the mean change score over the three scales was 60, an *S* scoring higher than 60 was above the average in over-all susceptibility to change. Two groups of *S*s were formed using only these total standard scores: (*a*) *susceptible* group, *S*s having standard scores of 67 or more; (*b*) *negative change* group, *S*s having total standard scores of 54 or less.[2]

*S*s whose total scores ranged close to the mean change were divided into two groups according to the variability of their reactions among the three scales. On the basis of a variability score, yielded by the sum of the differences between the *S*'s standard score on each scale and the corresponding mean standard score, the third and fourth groups were established as follows: (*c*) *little change* group, *S*s having total standard scores of 55 to 66 and variability scores of 7 or less; (*d*) *variable change* group, *S*s having total standard scores of 55 to 66 and variability scores of 8 or more. The scores defining the four groups were selected so that the *N*s in the groups would be nearly equal.

To study the personality characteristics associated with the four kinds of reactions to the communication, the four attitude change groups were used in setting up analysis of variance designs with cognitive complexity-simplicity, perceived similarity to peers, and perceived similarity to authority as dependent variables in the analyses. If attitude change had been taken as the dependent variable for purposes of the analysis, these kinds of reactions to outside influence could not readily have been differentiated. An *N* of 63 in the complexity analysis was reduced to 61 in the assumed similarity measures because two *S*s had not recorded any marks in the self column of the RCRT. Because of disproportional cell frequencies, the analysis of variance procedure suggested by Snedecor (9) was employed.

TABLE 2

ASSIMILATIVE PROJECTION ON PEERS IN RELATION TO SOURCE OF INFLUENCE AND TYPE OF ATTITUDE CHANGE

Type of Attitude Change	Mean Assimilative Projection on Peers	
	Generals (*N*) as source	Peers (*N*) as source
Little change	(9) 19.56	(6) 24.50
Variable change	(5) 20.80	(9) 21.89
Susceptible	(6) 20.50	(8) 23.75
Negative change	(6) 24.17	(12) 19.50

Analysis of Variance

Source of Variation	*DF*	*MS*	*F*	*p*
Source of influence	1	11.18	—	—
Type of attitude change	3	5.21	—	—
Interaction	3	67.28	4.12	.05
Within groups	53	16.34		

TABLE 3

ASSIMILATIVE PROJECTION ON AUTHORITY FIGURES IN RELATION TO SOURCE OF INFLUENCE AND TYPE OF ATTITUDE CHANGE

Type of Attitude Change	Mean Assimilative Projection on Authority Figures	
	Generals (*N*) as source	Peers (*N*) as source
Little change	(9) 14.67	(6) 16.67
Variable change	(5) 13.60	(9) 15.67
Susceptibility	(6) 17.33	(8) 14.75
Negative change	(6) 12.67	(12) 12.92

Analysis of Variance

Source of Variation	*DF*	*MS*	*F*	*p*
Source of influence	1	.08	—	—
Type of attitude change	3	31.11	2.47	—
Interaction	3	16.23	1.29	—
Within	53	12.60		

[2] The standard scores were combined in order to obtain a measure of the individuals' general reactions to the external communication. The *S*s tended to be slightly influenced in the direction of the communications. The raw score change means on the three attitude scales were 0.0, 1.0, 2.0, and the *SD*s were 3.7, 3.5, and 6.0, respectively.

The *susceptible* *S*s generally changed in the direction of the communication. Twelve of the 16 *S*s changed toward the communication on all three scales; the other four were successfully influenced on two out of three scales. The *negative change* *S*s tended to change in the direction opposite to that advocated by the communication. Sixteen of the 18 *S*s changed away from the communication on two or three of the three scales.

RESULTS AND DISCUSSION

Hypothesis I states that attitude change is less likely to occur in "cognitively simple" persons. Since the "little change" group contains persons who maintain their attitudes in the face of the outside influence, this group should have the greatest cognitive simplicity. The results of the analysis of variance given in Table 1 indicate that there are significant differences among the attitude change conditions when the communication source is not considered. The little change group has the highest cognitive simplicity means, tending to support the hypothesis.

The mean cognitive simplicity in the "little change to peers" condition is significantly greater ($p < .05$) than the mean in the "little change to generals" condition, suggesting that the hypothesized relationship between little change in the attitude scale and cognitive simplicity is most pronounced when the source of the attitude influence is the *Ss*' peer group. It may be that a "cognitively simple" person is not as resistant to change when the source is authority figures because of a need to conform to these authority figures. Interestingly, the groups with the highest cognitive complexity are those negative in their reactions to both peer and authority persuasion.

Hypothesis II states that persons who are susceptible to peer influence are more likely to perceive peers as similar to self than persons who are negative to peer influence or are susceptible to authority influence. This hypothesis is supported by the significant interaction ($p < .05$) in the analysis of variance presented in Table 2. The *Ss* in the "susceptible to peers" condition perceive themselves more similar to peers than the *Ss* in the "negative to peers" condition ($p < .05$), while the direction of the differences between the susceptible and negative conditions is reversed when generals are the source of influence. Although this latter difference does not attain significance, it may be that *Ss* who perceive themselves similar to peers react negatively to opinions advanced by generals.

The difference in "assimilative projection on peers" between the *Ss* susceptible to peer influence and those susceptible to the generals is in the predicted direction, although it fails to reach significance. The high "mean assimilative projection on peers" in the "little change to peers" condition was not predicted.

Table 3 presents the mean scores for similarity to the authority figures in the eight groups. Analysis shows no significant effects, and Hypothesis III regarding similarity to authority is therefore not supported. Perhaps "authority" was considered too broadly to be psychologically meaningful. Reaction to authority may be different according to the kind of authority; for instance, conveyers of society's values, charismatic leaders, or parental figures.

SUMMARY

Students in two undergraduate psychology courses were administered the same attitude scales on two occasions. The second administration was accompanied by information concerning the expressed opinions of either generals or peers. As hypothesized, those students showing little change in their attitude scale scores were found to be high on a measure of cognitive simplicity. Those students who were influenced by peers perceived themselves as more similar to peers than students who were negative to peer influence. An hypothesis relating assimilative projection on authority figures to influence by authority was not supported by the data.

REFERENCES

1. BERKOWITZ, L., & LUNDY, R. M. Personality characteristics related to susceptibility to influence by peers or authority figures. *J. Pers.*, 1957, **25**, 306–316.
2. BIERI, J. Cognitive complexity-simplicity and predictive behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 263–268.
3. FIEDLER, F. E. The psychological-distance dimension in interpersonal relations. *J. Pers.*, 1953, **22**, 142–150.
4. HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion*. New Haven: Yale Univer. Press, 1953.
5. JANIS, I. L. Personality correlates of susceptibility to persuasion. *J. Pers.*, 1954, **22**, 504–518.
6. KELLY, G. A. *The psychology of personal constructs*. New York: Norton, 1955.
7. LEWIS, HELEN BLOCK. An experiment on the operation of prestige suggestion. In G. E. Swansen, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Henry Holt, 1952. Pp. 18–29.
8. LUNDY, R. M. Assimilative projection and accuracy of prediction in interpersonal perceptions. *J. abnorm. soc. Psychol.*, 1956, **52**, 33–38.
9. SNEDECOR, G. W. *Statistical methods*. (4th ed.) Ames, Iowa: Collegiate Press, 1946.

Received August 1, 1956.

MAKE-A-PICTURE STORY (MAPS) TEST FINDINGS FOR RUBELLA DEAF CHILDREN[1]

D. MARJORIE BINDON

Canterbury University College, New Zealand

ALTHOUGH the Psychological Abstracts (down to 1955) do not report any personality studies done specifically on rubella deaf children (children born deaf through the mother having German measles during pregnancy), many people closely connected with the deaf are of the opinion that these children behave differently from other defective hearing children. The writer, as a trained teacher of the deaf, has both taught and lived with hearing-impaired children and her experience has led her to query this belief. It is her view that rubella deaf children and nonrubella deaf children (those born deaf through causes other than maternal rubella or who had become deaf before the age of two) seemingly display similar personality characteristics but that, as a group, defective hearing children who have not naturally acquired speech and language do differ from the hearing children.

To test this hypothesis the writer (2, 3) studied the personality characteristics of rubella deaf children by means of the Rorschach test using the Munroe checklist system of scoring (7). Her findings confirmed her hypothesis and revealed a common rigid personality pattern differentiating both the rubella deaf and the nonrubella deaf from the hearing group at well beyond the .001 level of confidence.

This article reports findings with the MAPS test[2], also used in the same investigation. The MAPS test can be adapted to avoid a frequent limitation of personality studies of deaf children that do not take fully into account the considerable language retardation of the deaf. When the examiner is a trained worker with hearing impaired children, and when the procedure is altered to eliminate the necessity for verbal response and the data are limited strictly to the scoring of the formal signs, the MAPS technique can seemingly be made entirely objective, quantitative, and if necessary, nonverbal in performance.

Perhaps the most serious handicap of deafness lies in the fact that it prevents the establishment of normal interpersonal relationships (cf. 1, 4, 8, 9). The fantasy productions of deaf children may also be expected to reflect this deficit. The general hypothesis as previously stated, therefore, leads to the following specific prediction to be tested here: MAPS records of rubella deaf and nonrubella deaf do not differ in the number of "normal" signs and "schizophrenic" signs; but the rubella deaf score fewer "normal" signs and more "schizophrenic" signs than their hearing counterparts.

METHOD

Subjects

Rubella deaf group. This group was comprised of 36 fifteen-year-old rubella deaf children (16 boys, 20 girls) who were born between December 1, 1938, and July 31, 1939. When tested, they were attending coeducational government schools for the deaf in New Zealand which use entirely oral methods and admit both day and residential pupils. All the qualified teachers in these schools are trained at the main school for the deaf and follow similar educational procedures. The children studied[3] represent a rubella deaf population that lies within the extremes of intellectual functioning, since both the ineducable, or those with serious multiple handicaps, and the most intellectually able were not, at the age of 15, attending a school for the deaf.

To permit investigating whether sex, degree of deafness, or the residential factor influence test performance differentially, this group was chosen to include as many nonresidential pupils as possible, and an equal number of "partially deaf" and "totally deaf"

[1] Based on part of thesis in partial fulfillment of requirements for the degree of Master of Arts in Psychology, University of New Zealand.

[2] MAPS test (10) is a variation of the TAT, but is less structured as figures are separate from the backgrounds. *S* selects and places figures on each background picture in turn and then proceeds to tell his story for that particular card. Test material comprises 22 background pictures $8\frac{1}{2}'' \times 11''$, from which *E* normally selects 11, and a set of 67 figures identifiable by a code symbol from a figure identification card. Data derived from figure location sheet can be subjected to objective quantitative treatment whereas *S*'s stories can be analysed in the manner of TAT protocols.

[3] The names of the children whose deafness was due to maternal rubella, or to other known causes before the age of two, were compiled from the original data of D. R. Hay (5).

children who could be matched for age, sex, and educational classification (above average, average, below average). "Partially deaf" and "totally deaf" refer here, respectively, to those children with average hearing losses >60 decibels but <101 decibels, and >100 decibels in the better ear, as tested by a pure-tone audiometer in the frequencies 512, 1024, 2048. The partially deaf children could use their residual hearing to varying extents with the help of individual hearing aids. Hence, they acquired language at a faster pace and, in general, were taught separately from the severely hearing impaired of similar age and intellectual ability.

Nonrubella deaf group. This group was comprised of 15 fifteen-year-old children (7 boys, 8 girls) who were either born deaf through causes other than maternal rubella or who had become deaf before they were two, and were all attending the government schools for the deaf in New Zealand. The writer is not prepared to say that these children are truly representative of a nonrubella deaf population; they were the only nonrubella deaf children available whose etiology had been clearly determined.[4]

Nondeaf group. This group was comprised of thirty 15-year-old children (13 boys, 17 girls) selected from children who volunteered to be tested and were attending a coeducational, multicourse, public, nonresidential high school that was situated in the same city as the main school for the deaf. The school was selected to represent a hearing city school population that fell within the extremes of the very highest and lowest intellectual levels and whose members were reasonably well distributed throughout the various socioeconomic categories as gauged by parental occupation. A public day school was chosen rather than a private residential school or an orphanage, since the latter types of institution not only do not sample a cross section of the school population, but are also far more organized, formal, and restrictive in their supervision of the out-of-school activities of their charges than are the government residential schools for the deaf, which are run on very progressive lines.

Because of the limited number of deaf subjects available for testing, the method of matched pairs was used to investigate intergroup and deaf subgroup differences. Therefore, from the rubella deaf group of 36 children, the 30 who were most closely matched for age, sex, educational classification, and approximate socioeconomic status with a member from the volunteer normal hearing control group were selected for comparisons with hearing children.[5] Similarly, but with the additional criterion of degree of hearing loss, 15 children from the rubella deaf group were paired with children in the nonrubella deaf group. Further, to investigate the possible influence of the residential factor on test performance, each of the nonresidential deaf pupils (7 rubella deaf, 5 nonrubella deaf) was also matched for sex, age, degree of deafness, educational classification and approximate socioeconomic status with an appropriate member from the residential deaf group.

In conjunction with the Rorschach test and the MAPS test, the Wechsler-Bellevue Intelligence Scale Performance Form 11 was administered to all subjects (*Ss*) in this study. Statistical analysis of results by means of the *t* technique for matched pairs revealed no significant differences between the compared groups or deaf subgroups in either mean IQ or mean subtest scores, the pattern of which was essentially similar ($p > .1$ in all cases). It should be noted, however, that these mean scores for the rubella deaf group, as for the nonrubella deaf group, were all consistently lower than for the nondeaf group; that these findings were not significant may have been due to the small size of the samples. (Mean IQ of rubella deaf group was 91.6, and *SD* 12.2, as against mean IQ 101.7, and *SD* 13.6 for the nondeaf group.)

Administration

To overcome any bias the MAPS test may have against deaf children who are retarded in language development, the experimenter (*E*) altered the test procedure so that *Ss* could respond nonverbally to each stimulus situation. In the modification, *Ss* are not asked to relate a complete story about the figures they have selected to people each background picture. Content analysis in the manner of TAT interpretation is therefore precluded, and the scoring and analysis of the test findings is limited to the formal psychosocial sign categories that Schneidman (11) found to differentiate the fantasy productions of normals from psychotics[6]. The test was given individually to each *S* in a room occupied only by him and *E*. It came as the last of the three tests required of him. The specific procedure used for all *Ss*, nondeaf as well as deaf, was as follows:

The *E* emptied the box containing the cutout figures onto the table at which *S* was seated. The *S* rearranged the figures face upwards. *E* sat slightly behind and to the left of *S*. From his particular position not only could *E* record responses without distracting *S*, but also a deaf *S* could speech read *E* with relative ease. The background pictures: Livingroom, Street Scene, Medical, Bathroom, Dream, Bridge, Bedroom, Blank, Doorway, Cemetery, and Forest, were presented in that order. *E* issued the following instructions orally:

> Have a good look at all the little pictures on the table. There is no hurry. Tell me when you are finished. . . . Now here is a big picture (first background picture placed on table in front of *S*). I want you to make me an interesting story by choosing pictures from any of these here (indicating

[4] Ibid.

[5] The composition of the nondeaf group in respect to age, sex, educational classification, and approximate socioeconomic status was directly comparable to that of the rubella deaf group. The writer had intended originally to match each rubella deaf child for the above mentioned variables with a nondeaf child, but she was forced to select her normal control group from volunteers of whom only 30 fulfilled her matching criteria.

[6] Shneidman (11) established a set of 59 objective signs (39 "normal" and 20 "schizophrenic") that could statistically differentiate these two groups. Shneidman defines a MAPS test sign as "a denotable aspect of the subject's performance, specifically having to do with his selection and handling of the test material." (11, p. 189). It refers to a structural aspect of the performance as against the content of performance and relates to what is done with the figures rather than what is said about them.

figures on table) and putting them on this picture (pointing to background picture). You can use as many of these pictures (pointing to figures) as you like. Off you go.

If *S* put just one figure on the background picture he was told, for the first card only, that he could use more than one figure if he wished. This instruction was given in case a deaf *S* had misunderstood the opening test instructions. *E* then told *S* "Yes, that's right. Now tell me about the story." (*E* was interested only in gaining information to help her to record the theme and decide whether figure interaction had taken place and was not concerned with securing content material for a possible TAT-like analysis: verbal responses were not essential for securing this information which could as readily have been obtained pantomime fashion.)

Well, that was a good story. I have some more pictures here (indicating pile of background pictures) for you to tell me a story about. You can use the same little pictures here (pointing to figures on first background picture) again, or you can use any of the others (indicating remaining figures with a sweep of the hand).

The *S* then returned the figures he used to the table, the first background picture was withdrawn and another put in its place. All the background pictures were dealt with in the same manner. Schneidman was followed explicitly in the formal scoring of the figure location sheets.

RESULTS AND DISCUSSION

Quantitative Findings

To avoid capitalizing on chance differences (cf. 6, p. 245), it was decided not to compare groups on their total individual sign scores unless their composite sign score differences yielded a significant $\chi^2 p < .01$ (χ^2 corrected for continuity, $df = 1$, cutting point lower median score).

Both MAPS composite sign scores, "normal" and "schizophrenic," failed to differentiate the nonrubella deaf group from its matched rubella deaf group; they also failed to differentiate the matched deaf subgroups formed on the basis of degree of deafness, sex, and residential factor, respectively. (The differences in all cases were not large enough to warrant statistical analysis.) These results are consistent with the first half of the hypothesis under investigation; namely, that rubella deaf children do not differ significantly from nonrubella deaf children in number of "normal" signs and "schizophrenic" signs on the MAPS test. Moreover, neither sex, degree of deafness, nor the residential factor appeared to have had any differential influence on the fantasy productions of the deaf population tested as measured by these composite sign scores.

On the other hand, in confirmation of the latter half of the specific prediction, the MAPS test results show that the rubella deaf sample gave significantly fewer "normal" signs ($p < .001$) and significantly more "schizophrenic" signs ($p < .01$) than did its matched hearing group (Table 1). Because these over-all results were significant, the scores for these groups were examined for differences in the separate sign categories. Out of a total of 59 sign comparisons, 6 "normal" signs and 3 "schizophrenic" signs were found to differentiate the above groups ($p < .05$—Table 1).

The deaf (rubella *and* nonrubella) and nondeaf *S*s were then compared on these particular 9 signs and on their composite sign scores, the last child in the rubella deaf sample of 36 being arbitrarily dropped so as to facilitate statistical calculations. This step was taken in order to test the general hypothesis that, *as a group*, defective hearing children who have not naturally acquired speech and language behave differently from the hearing. The statistical findings confirmed this hypothesis. The sign category scores on which the deaf and hearing subjects were compared were, without exception, found to differentiate them at beyond the .05 level of confidence; both composite sign scores and all individual sign categories, except for *FI* (figure interaction) with spouse and the presence of *M-8* (priest) in any background (both dropped from $p < .02$ to $p < .05$) had chi squares that were larger than those calculated for the nondeaf and the paired rubella deaf group (Table 1).

Qualitative Findings

Only qualitative information that supplements the quantitatively established differences will be discussed here. Since the qualitative performances characteristic of the rubella deaf are also characteristic of the nonrubella deaf, the discussion in this section refers to the deaf as a group.

The outstanding fact to emerge from comparison of deaf and nondeaf on the "normal" sign scores was that the group difference in each figure interaction category was highest when the interaction represented the most personal and frequent interrelationships of normal everyday life, and was lowest for the least personal and least frequent. Thus, *FI* with child by adult gave $p. < 001$; *FI* with

TABLE 1

COMPARISON OF MATCHED RUBELLA DEAF AND NONDEAF, AND OF DEAF AND NONDEAF GROUPS WITH RESPECT TO CERTAIN MAPS SIGNS*

Category	Rubella Deaf ($N = 30$)	Non-deaf ($N = 30$)	χ^2	p	Deaf ($N = 50$)	Non-deaf ($N = 30$)	χ^2	p
"Normal" Signs								
"Normal" composite sign score	12	28	16.86	$<.001$	22	28	17.42	$<.001$
Child figure three or more times	2	9	4.01	$<.05$	5	9	14.9	$<.001$
Figure interaction with acquaintance	19	27	4.57	$<.05$	31	27	6.04	$<.02$
Figure interaction with spouse	12	22	5.50	$<.02$	22	22	5.39	$<.05$
Figure interaction with parent	5	23	17.87	$<.001$	7	23	28.8	$<.001$
Figure interaction with child by adult	3	21	20.07	$<.001$	6	21	25.75	$<.001$
Ir· Medical, man takes an exam	1	11	8.44	$<.01$	1	11	15.06	$<.001$
"Schizophrenic" Signs								
"Schizophrenic" composite sign score	22	11	6.73	$<.01$	38	11	9.98	$<.01$
Feet off ground	17	5	8.68	$<.01$	32	5	15.48	$<.001$
In any background—priest, *M-8*	19	9	5.42	$<.02$	28	9	4.11	$<.05$
Ignoring background	18	1	19.72	$<.001$	30	1	23.04	$<.001$

* In testing "normal" and "schizophrenic" composite sign scores for significant differences, lower median score for matched rubella and nondeaf groups was used as cutting point (9 and 1, respectively).

parent $p < .001$, *FI* with acquaintance $p < .02$, *FI* with spouse $p < .05$; in comparison with the remaining *FI* categories where no statistically significant differences existed: *FI* with stranger and *FI* with animal gave $p > .1$ in both cases. The lack of themes involving figure interaction on the part of the hearing-impaired was not so much due to their giving more stories to backgrounds that contained only one figure, but to their tendency to populate the backgrounds with figures each going his own way or, apparently, representing separate stories. For instance, a street scene of six figures with no interaction between any of them may be given: or, in the dream, four figures may be placed, each of whom represents a different dream that the sleeper had had.

Group differences in the selection of the same child figure three or more times ($p < .001$) may indicate that the defective hearing group show less constancy in their identification with a child figure. This indication of a lack of constancy in identification seems to be supported by the investigator's observation that the individual deaf *Ss*, more than the hearing, tended to use a greater variety of figures to represent the same type of character (e.g., a father or mother); in doing this they tended, more than did the nondeaf, to include in their selection of human figures, figures outside the male adult, female adult, and child categories (e.g., minority category: Negroes, Jews, etc., silhouette and blank faces category; and undetermined as to sex category).

Although the nondeaf sample gave a higher proportion of figure relationships involving verbal interaction, the nature of the interrelationships was, in general, very similar for both groups: Physical aggression (mainly shooting or fighting) predominated, and the most common feelings expressed were anger, fear, and to a lesser extent, sorrow, while the emotions of love, joy, or happiness were conspicuously absent.

It has already been mentioned that an outstanding characteristic of the fantasy productions of the deaf is their lack of themes involving interpersonal relationships; another outstanding feature is their element of incongruity due, apparently, to a lack of reality testing (cf. "schizophrenic" signs: ignoring background, $p < .001$; feet off ground, $p < .001$). Information from the figure location sheets supplied other types of inappropriateness given by the deaf but not by the hearing group. For example, specific figures were selected for specific backgrounds for which they are not normally selected: *F-1* [nude female figure] in the Doorway, *L-4* [ghost] in bath in the Bathroom; incongruous themes were given: "This man [*M-6*—man with a gun] shot that man [*M-7*—supine blood-stained male] and policeman [*M-5*—policeman standing feet astride, hands on hips] saw"; the

placing of *M-7* on the operating table in the Medical and a male figure beside it, with the theme being that *M-7* was dead and the other figure was the doctor. It is of interest to note that these two themes, or very similar variations of them, were given by several deaf children as against none of the hearing, and it is tentatively suggested that they exemplify not only incongruity, but also the assigning of ineffectual roles to people in authority.

The more frequent use that the deaf made of the schizophrenic sign *M-8* (priest, used almost exclusively in the Cemetery background) can probably be accounted for in terms of an association of ideas and does not in their case indicate a heightened religiosity (cf. Shneidman, 11).

SUMMARY

With procedure altered slightly to eliminate the deaf subjects' language handicap, the MAPS test was given to three groups of 15-year-old children: 36 rubella deaf, 15 nonrubella deaf, and 30 hearing children. For purposes of intergroup matched pair comparisons, members of the nonrubella deaf group and nondeaf group, respectively, were matched with individuals selected from the rubella deaf group; the groups to be compared being then (*a*) rubella and nonrubella deaf (15 pairs) and (*b*) rubella deaf and nondeaf (30 pairs).

Results (quantitative and qualitative) for both deaf groups were similar and when matched deaf subgroups were compared, neither degree of deafness, sex, nor institutional residence was found to have had any specific influence on their performances ($p > .1$ in all cases). However, the rubella deaf, in comparison with the nondeaf, scored fewer "normal" signs ($p < .001$) and more "schizophrenic" signs ($p < .01$). Six "normal" signs and three "schizophrenic" signs were found to differentiate these two groups ($p < .05$).

Furthermore, a comparison of the deaf as a whole (rubella *and* nonrubella, $N = 50$) with the nondeaf revealed that the above signs could also distinguish these two populations ($p < .05$ in all instances).

It is contended that the fantasy productions of the deaf, in general, are indicative of their social isolation and illogical unrealistic thinking.

REFERENCES

1. BEST, H. *Deafness and the deaf in the United States*. New York: Macmillan, 1943.
2. BINDON, D. MARJORIE. Personality characteristics of Rubella deaf children: Implications for teaching of the deaf in general. *Amer. Ann. Deaf.*, 1957, **102**, 264–270.
3. BINDON, D. MARJORIE. Rubella deaf children: Rorschach study employing Munroe Inspection Technique. *Br. J. Psychol.*, in press.
4. GETZ, S. *Environment and the deaf child*. California: California School for Deaf, 1955.
5. HAY, D. R. Relation of maternal rubella to congenital deafness. *N. Z. Med. Journal*, 1948, **48**, 604.
6. MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.
7. MUNROE, RUTH L. The inspection technique for the Rorschach protocol. In L. E. Abt, & L. Bellak (Eds.), *Projective psychology*. New York: A. A. Knopf, 1950. Part 11.
8. MYKLEBUST, H. Clinical diagnosis of children with impaired hearing. *Volta Review*, 1946, **48**, 738–742.
9. ADVISORY COUNCIL ON EDUCATION IN SCOTLAND. A report on pupils who are defective in hearing. Edinburgh: His Majesty's Stationery Office, 1950.
10. SHNEIDMAN, E. S. A manual for the MAPS test. *Proj. Tech. Monogr.*, 1951, **1**, No. 2.
11. SHNEIDMAN, E. S. Schizophrenia and the MAPS test. *Genetic Psychol. Monogr.*, 1948, **38**, 145–233.

Received August 7, 1956.

SHIFT IN JUDGMENT OF WEIGHTS AS A FUNCTION OF ANCHORING STIMULI AND INSTRUCTIONS IN EARLY SCHIZOPHRENICS AND NORMALS[1]

KURT SALZINGER[2]

Columbia University

THE literature concerned with investigation of the variables which are effective in changing an absolute scale of judgment is a rich one (1, 3, 5, 6, 7, 9, 10, 13, 15, 16) dealing with many varied types of stimuli which impinge upon the different sense receptors of the organism. The consistency of findings has led investigators to postulate a single judgment process to explain the formation of and shift in a scale of absolute judgment with normal subjects (*S*s).

In this study, differences in the judgment mechanism between normal *S*s and schizophrenics were studied by using a modified form of the method of absolute judgment. The modification was employed because a pretest of the method indicated that in the presence of the anchor stimuli some *S*s instructed themselves to shift their judgments and some to keep them constant. The instruction variable was therefore controlled by telling *S*s not to shift. This change in the administration of the method of absolute judgment made it possible to describe *S*s' judgments in the anchor condition in terms of two conflicting discriminative stimuli. The first of these, a physical stimulus, was the anchor in conjunction with the test stimulus following it. The second, a verbal stimulus, consisted of the instructions of experimenter (*E*).

In terms of Goldstein's conceptualization of schizophrenic thinking (2), reaction to the verbal stimulus may be viewed as an instance of abstract thought, whereas reaction to the physical stimulus characterizes concrete thought. Amount of shift (i.e., the change in judgment from the unanchored to the anchor condition) can then be said to vary as a function of the extent to which a given *S* reacts to concrete as contrasted to abstract stimuli.

The shift phenomenon may be looked upon from another point of view, that of constancy. If the instruction variable is effective, then it serves to keep the judgment constant. Whether or not an anchor stimulus precedes it, the same judgment is evoked for a given stimulus. The extent to which *S* shifts then serves as an index of his inability to preserve constancy. In many instances, constancy is maintained when the response to abstract aspects of the stimulus or stimulus situation is prepotent over the response to the concrete aspects of the stimulus or stimulus situation.

On the basis of this formulation, schizophrenics and normals were compared on their ability to respond to concrete versus abstract stimuli, or on their ability to maintain constancy. The central hypothesis was that schizophrenics are less able to maintain constancy or to react to abstract (verbal) stimuli than normals.

Other studies have been made using similar techniques on patient groups. Sonder (14) found judgment of lifted weights to be prognostic in a study of schizophrenics. Kaplan and Helson (4) demonstrated differences in performance between lobotomized patients and normals in a pilot study on a weight judgment task. Zubin (18) used the same weight judgment technique as that employed in this study for a comparison of patients who leave the hospital with those who do not. He found that patients who stayed in the hospital shifted more than those who left.

Rado's (11) theory that proprioceptive disorder is an innate trait of schizotypal organization gave rise to the second hypothesis of this study; namely, that schizophrenics make more incorrect judgments than normals in a task where judgments depend upon proprioceptive discrimination.

[1] This investigation was supported in part by research grant M586C from the National Institute of Mental Health. The article is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree in the Department of Psychology in the Faculty of Pure Science, Columbia University in 1954. The author is indebted to Professors J. Zubin, H. Garrett, C. G. Mueller, Dr. S. Kugelmass, and Mrs. S. Salzinger for their invaluable aid during the various phases of the dissertation on which this paper is based.

[2] Now at Biometrics Research, N. Y. State Dept. of Mental Hygiene, Psychiatric Institute, 722 West 168 Street, New York 32, N. Y.

Mueller and Schumann (8) indicated that the lighter a weight, the higher and faster it is lifted, and that judgment of heaviness depends upon *S*'s response to the ease with which he overcomes each weight's inertia. The design of this experiment allowed a retest of this contention as the third hypothesis.

Both judgment time and height of lift of each weight were further explored as relevant parameters of the judgment situation. These variables were examined in order to elucidate any differences in shift found between normals and patients.

METHOD

Apparatus

The weight judgment apparatus consisted of a turntable which could be revolved in such a way as to make available one stimulus at a time to *S*, an arm rest allowing *S* to keep his hand in a comfortable and steady position, and seven weights of 100, 200, 250, 300, 350, 400, and 900 gms. Each of the weights was encircled with lines marking one-cm. distances from the bottom so that there was a total of seven such lines. A buzzer-chronoscope circuit was so set up that the buzzer served as a signal for *S* to lift the weight, and the time from the signal to the time of judgment could be obtained as an index of judgment time.

Subjects

Sixteen acute schizophrenic patients (eight males and eight females) were randomly chosen from one male and one female ward of the New York State Psychiatric Institute. All patients were cooperative and in good contact. Sixteen normals (eight males and eight females), most of whom were chosen from the School of General Studies of Columbia University, were equated with patients (matched by pairs) as far as was practically feasible in terms of sex and age. The mean age of the patient group was 27.6 years (range from 19 to 38 years), while that of the normal group was 27.2 years (range from 19 to 39 years). Patients ranged in educational status from one term in high school to the Ph.D., while the education of the normals ranged from one year of college to the Ph.D. The diagnoses of the patients were: 3 schizophrenic undifferentiated, 6 schizophrenic paranoid, 2 schizophrenic mixed, 3 schizophrenic catatonic, 1 schizophrenic simplex, and 1 schizophrenic pseudoneurotic. Patients participated in the experiment as part of their regular routine while in residence in the hospital. Normals were paid for their participation in the experiment. Unfortunately, two normal *S*s were presented with the experimental conditions in incorrect order and had, therefore, to be discarded from the sample in the second session. When normal patient comparisons were made, the patients matched with the discarded normals were left out of the analysis.

Experimental Procedures

Both normals and schizophrenics underwent the same experimental procedures. The *E* ascertained from each *S* whether he engaged in an occupation involving weight estimation (no such *S* was found) and whether he was right or left handed (*S* was subsequently requested to use his dominant hand). All *S*s were presented with the following instructions upon entering the experimental room:

> This is a weight judgment task. I am going to ask you to judge the heaviness of five weights in comparison to each other by calling the heaviest "very heavy," the next heaviest "heavy," the next one "medium,"—that's the middle weight—the next one "light," and the lightest one "very light."
>
> Now suppose you place your arm in this arm rest (right for right-handed, left for left-handed). Right underneath your hand you will find the knob of a weight. When I give you the signal like this (sound buzzer), you will lift the weight approximately one inch from the table top, set it down, and give me your judgment of it. Now just to show you the correct judgments, I'll present the weights to you in order, telling you which judgment applies to each of the weights. This is the very heavy weight (sound buzzer), this is the heavy one (buzzer), this is the medium one (buzzer), this is the light one (buzzer), and this is the very light one (buzzer).

This procedure familiarized *S* with the proper categories and trained *S* to respond to the buzzer signal and to lift the weight the required distance by means of his thumb and two neighboring fingers. After this initial training period, *E* made no additional comments concerning *S*'s height of lift. At the end of the training period *E* continued:

> O.K.? Now you have lifted all the five weights. From now on I'll present the weights to you in a mixed up order and you'll give me your judgment of the heaviness of the weight in terms of very heavy, heavy, medium, light, and very light. O.K.? I'll sound the buzzer, you lift the weight, set it down and give me your judgment of it.

The *E* then presented *S* with the five stimulus weights (200, 250, 300, 350, and 400 gm.) five times each in the pattern of a Latin square. Each weight was thus presented in a different position in each series of five weights. This unanchored condition was presented to each *S* at the beginning of the experiment in exactly the same way. Four male and four female normals and their matched patient *S*s were then presented with a heavy anchor; the rest of the *S*s were presented with a light anchor. The instructions for this anchor condition were:

> From now on I'm going to ask you to lift an extremely heavy (light) weight before lifting each of the weights that you are to judge. Don't bother to judge the first weight and make sure that its heaviness does not affect your judgment of the second weight. O.K.?

In this condition, the weights were presented to *S* in exactly the same order as in the unanchored condition, except that *S* was required to lift the anchor before each weight that he judged.

At the end of the anchored session, *S* was allowed to remove his arm from the arm rest, and *E* answered questions (without disclosing the purpose of the experiment) for a period of approximately five minutes. When the rest period ended, *E* trained *S* as in the beginning of the experiment, repeated the instructions,

and then presented *S* with the unanchored condition. At the end of this period, those *S*s who had been presented with the heavy anchor before the rest period were presented with the light anchor, while those who had been presented with the light anchor before the rest period were presented with the heavy anchor. This second anchored condition was presented to *S* with the same instructions as before.

Fourteen patients and 11 normals were retested after a lapse of two days; the other 2 patients were retested after 4 days, while 2 of the normals were retested after 3 days, 1 after 1 day, 1 after 4 days, and 1 after 8 days. All *S*s were retested with the same procedure except that the order of the anchor conditions was reversed. The *S*s who had previously been given the heavy anchor first were now given it second, and vice versa.

The test weights in the unanchored condition and the anchor weights in the anchored condition were presented to each *S* approximately eight seconds after the preceding judgment; the test weights in the anchored condition were presented approximately four seconds after *S* set the anchor weight down.

During the eight-second period, *E* recorded the judgment of *S*, the height of lift correct to the nearest centimeter, and the decision time correct to the nearest .01 of a second.

The decision time as measured included the time between the sounding of the buzzer and *S*'s lifting of the weight, *E*'s reaction time to *S*'s vocal judgment (disconnecting the chronoscope circuit), and *S*'s decision time proper (the time from lifting the weight to the judgment). Since the reaction times involved were much shorter than the whole decision time, however, the latter could still be used for comparative purposes as it was in this experiment.

None of the *S*s was aware of the fact that *E* was recording either their decision time or their heights of lift.

The raw data were converted to stimulus shift scores and kinaesthetic scores as described in detail by Salzinger (12). These scores were based on the difference in grams between the weight to which a particular judgment was assigned and the one to which it should have been assigned according to the verbal instructions given during the preliminary training period.

RESULTS

Shift

Combining the shift scores of both test and retest sessions, one shift score was obtained for the heavy anchor and one for the light anchor for each *S*. A comparison of normals and schizophrenics using Wilcoxon's paired replicates test (17) showed that the latter shifted significantly more than the former ($.02 < p < .05$) in the heavy anchor condition but did not differ from the normals in the light anchor condition ($p > .05$).

For the first session alone, the results were the same. The schizophrenics shifted more than the normals ($.01 < p < .02$) from the unanchored to the heavy anchor condition, but no difference was found from the unanchored to the light anchor condition ($p > .05$). For the second session alone, there were no significant differences between normals and patients for either anchor.

Since there was a significant difference in shift between normals and patients in the first session but not in the second, the changes from the first to the second session were examined. Wilcoxon's paired replicates test demonstrated a significant ($p = .05$) increase in shift due to the heavy anchor for normals from the first to the second session. There was a nonsignificant trend in the opposite direction for the patients. No significant differences in shift were found for the light anchor conditions. The contrary changes in both groups were apparently responsible for the elimination of the significant difference between them.

Change in relative status of each *S* from the first to the second session was examined in terms of rank order correlation coefficients between shift scores. Statistically significant correlations ($p < .05$ by a one tail test) were found for the shift due to the heavy anchor of both normals ($N = 14$, rho $= +.49$) and patients ($N = 16$, rho $= +.44$). On the other hand, correlations for the light anchor condition for both normals (rho $= +.11$) and patients (rho $= +.33$) showed no such relationship.

Kinaesthesis

There was no statistically significant difference ($p > .05$) by Wilcoxon's (17) paired replicates test between normals and patients in their kinaesthetic ability. This was tested by comparing the kinaesthetic score (total error score in grams) of the patients and their controls on all unanchored conditions. Comparison of the change in discriminative ability from the unanchored to the anchored conditions in normals with that of the patients also revealed no statistically significant difference.

Decision Time

Decision time data were tabulated separately for each session, for each weight, and for each condition (unanchored, heavy anchor, and light anchor). Median values were then computed separately for the normals and patient. Results are shown in Table 1.

Wilcoxon's groups-of-replicates test indicated that the decision time of the patients

TABLE 1

MEDIAN DECISION TIME (IN SECONDS) OF JUDGMENTS OF NORMALS AND PATIENTS FOR FIVE DIFFERENT WEIGHTS UNDER CONDITIONS OF NO ANCHOR (NA), HEAVY ANCHOR (HA), AND LIGHT ANCHOR (LA) IN TWO SESSIONS

Sessions	Weights	Normals				Patients			
		NA	LA	NA	HA	NA	LA	NA	HA
1	200	3.90	4.44	3.86	4.13	4.06	4.56	4.36	3.83
	250	4.43	4.24	4.05	4.26	5.17	4.67	4.42	4.09
	300	4.42	4.32	3.88	4.30	5.36	4.81	4.88	4.79
	350	3.88	4.41	4.01	4.08	4.86	4.26	4.71	4.46
	400	3.94	3.76	3.84	4.19	4.82	4.13	4.46	4.35
2*	200	3.99	4.28	3.83	4.18	3.70	3.83	3.56	3.59
	250	4.61	4.37	4.51	4.59	3.71	3.86	3.89	4.01
	300	4.73	4.69	4.51	3.86	4.15	4.03	4.41	4.25
	350	4.38	4.46	4.47	5.01	3.86	3.98	4.34	4.18
	400	4.07	4.20	5.02	4.98	3.73	3.51	3.76	4.19

* There were 16 normals in the first session, 14 in the second session, and 16 patients in both sessions. Each *S* was presented each weight five times in each condition.

was significantly longer ($p < .01$) than that of the normals in the first session and significantly shorter ($p < .01$) than that of the normals in the second session.

Significant interaction, determined by a chi-square analogue of an analysis of variance in Wilcoxon (17), was found between groups (patient *vs.* normal) and experimental conditions (anchored *vs.* unanchored) in the first session only, i.e., the largest differences in decision time appeared in the unanchored condition. The group-condition interaction in the second session and the group-weight interaction in both the first and second sessions were nonsignificant.

Analysis of the change in decision time from the first to the second session showed a significant ($p < .01$) increase in decision time for normals and a significant ($p < .01$) decrease for patients by Wilcoxon's paired replicates test.

Session-condition interaction was significant ($p < .05$) for the patients, the heavy anchor condition showing the smallest change from the first to the second session; session-condition interaction for normals and session-weight interaction for both normals and patients were not significant.

Differences between conditions were tested for by a chi-square analogue of analysis of variance. A significant difference was found for patients between conditions in the first session ($p < .05$) but not in the second. No significant differences were found for either of the two sessions of normals. In the first session, patients showed the longest decision times in the two unanchored conditions, the shortest in the heavy anchor condition, and the next to the shortest in the light anchor condition.

Differences between weights were tested by the same technique and yielded no significant differences for either session of normals. For patients, significant differences were found between the decision times for different weights ($p < .05$ for first and second sessions). The longest decision time was associated with the middlemost weight in the series, the 300 gram weight, with increasingly shorter decision times for weights farther from the middlemost.

Height of Lift

The accuracy of *E*'s estimate of height of lift was determined by having a person lift a weight to a predetermined height by means of a pulley system and then recording *E*'s estimate of height. Ten per cent of the heights were incorrectly judged by *E*. No error exceeded .5 cm.

Height of lift data were tabulated separately for each session, for each weight, and for each condition (unanchored, heavy anchor, and light anchor). Median values were then computed separately for normals and patients. The results are presented in Table 2.

The height of lift data were subjected to the

TABLE 2

MEDIAN HEIGHTS OF LIFT (IN CM.) OF WEIGHTS OF NORMALS AND PATIENTS FOR FIVE DIFFERENT WEIGHTS AND UNDER CONDITIONS OF NO ANCHOR (NA), HEAVY ANCHOR (HA), AND LIGHT ANCHOR (LA) IN TWO SESSIONS

Sessions	Weights	Normals				Patients			
		NA	HA	NA	LA	NA	HA	NA	LA
1	200	3.3	2.0	2.9	3.5	4.3	3.5	4.0	4.0
	250	3.0	1.7	2.9	3.2	4.4	3.4	4.4	4.2
	300	2.9	1.8	2.5	3.2	4.3	3.5	4.2	4.2
	350	3.2	2.0	2.7	3.6	4.5	3.8	4.4	4.2
	400	2.8	2.1	3.2	3.0	4.2	3.6	4.4	4.1
2*	200	3.1	2.3	2.7	3.5	4.6	3.4	4.0	4.6
	250	2.6	2.3	2.5	3.2	4.4	3.6	4.3	4.5
	300	2.7	2.4	2.4	3.3	4.3	3.9	4.4	4.3
	350	3.3	2.6	2.6	3.4	4.2	3.7	4.4	4.4
	400	3.2	2.4	2.5	3.7	4.4	3.6	4.2	4.4

* There were 16 normals in the first session, 14 in the second, and 16 patients each in both the first and second sessions.

same kind of analysis as the decision time data. Wilcoxon's groups-of-replicates test indicated that the patients lifted their weights significantly ($p < .01$) higher than the normals in both sessions.

Group-condition interaction was significant at the .005 level for the first session and at the .03 level for the second session. In both sessions, the light anchor condition consistently produced the smallest differences between patients and normals. Group-weight interaction was not significant for either session ($p > .05$).

Height of lift did not change significantly ($p > .05$) from the first to the second session, nor was there a significant session-weight or session-condition interaction.

Analysis of differences between conditions and between weights by a chi-square analogue of analysis of variance (17), done separately for each session of normals and patients, yielded significant ($p < .05$) differences between conditions but not between weights. The heavy anchor condition consistently resulted in the lowest height of lift for both normals and patients. Among normals and in the second session of the patients, the light anchor condition resulted in the greatest height of lift, the unanchored condition falling in the middle between the light and heavy anchor condition. In the first session, the patients exhibited the second greatest height of lift in the light anchor condition and the greatest height of lift in the unanchored condition.

The relationship between the degree of shift and the change in decision time from the unanchored to the anchored conditions separately for normals and patients in each session and for each anchor was explored as a factor relevant to the judgment situation. Examination of the relationship yielded no significant rank-order correlations.

DISCUSSION

The central hypothesis that schizophrenics shift more than normals was confirmed for the heavy anchor condition. The ineffectiveness of the light anchor compared to the heavy anchor may perhaps be accounted for by its being only 100 gms. below the lightest test weight, whereas the heavy anchor was 500 gms. above the heaviest test weight.

The instability of the differential effect of the heavy anchor on normals and patients from test to retest required a more exact examination.

Analysis of the change in shift due to the heavy anchor showed that the lack of difference between normals and patients in the second session was primarily due to a change in the normals. They shifted significantly more in the second than in the first session, thus narrowing the initial large difference in shift between the two groups.

The correlations between test and retest showed that the shifts due to the light anchor were not reliable, while the shifts due to the heavy anchor for both patients and normals showed reasonable reliability ($p < .05$).

The results of the analysis of shift may be further clarified by exploring the relationship of decision time to shift. The change in decision time from the first to the second session paralleled that of shift; there was an increase in decision time for normals and a decrease for patients. Both these changes were significant. Further examination made it clear that differences between normals and patients were largest in the unanchored conditions. From these results it would appear that resistance to the anchor may be sustained only at the expense of a longer decision time than is otherwise necessary. This interpretation is consistent with the fact that in the first session, where the patients shifted more, their decision times were shorter for the anchored than for the unanchored conditions, while in the second session, where no significant differences in decision time occurred between the anchored and unanchored conditions, the patients shifted less. This notion applies to the first session for the normals, where there was no significant difference in decision time between anchored and unanchored conditions with relatively great resistance to shift, but does not seem applicable to the second session for normals, where slight resistance to shift was accompanied by nonsignificant differences between anchored and unanchored conditions. Furthermore, correlations of amount of shift with decision time within the group of normals and within the group of patients were not significant.

The height of lift data did not shed any light on the change in amount of shift from the first to the second session since neither

normals nor patients showed differential changes from one session to the other.

It now remains necessary to explain the instability in the normal-patient difference in ability to shift. Since only the normals changed significantly in their shift scores, it may be that their motivation decreased from the first session to the second. While they followed instructions closely in the first session because of their initial curiosity, they later became bored and had difficulty resisting the anchor; consequently, they took longer for their decisions. On the other hand, patients who showed decreased decision time, but no significant change in shift, may be supposed to have been as highly motivated to do well in both sessions but more anxious to get through with the task because of its potential threat (to them). This was evidenced by the fact that the patients showed concern during both testing sessions over whether their performance would have an effect on their remaining in the hospital in spite of all reassurances to the contrary by *E*.

Since the correlations between the shift scores of the two sessions were reliable though low the above interpretation is offered as a tentative explanation for the lack of difference in shift between normals and patients in the second session.

A comment might be made here concerning the low correlations. In general, the psychophysical techniques are used only after intensive training of highly motivated *S*s who have achieved high reliability in their judgments. When applying psychophysical techniques in an experiment involving patients, however, it becomes necessary to keep the task sufficiently short to hold the patients' attention. Repetition of the task on successive days, which in normals results in highly stable performance (learning has reached an asymptote), may encounter the typical fluctuations in the pathology of patients. It may be of interest to examine this problem by subjecting both patients and normals to longer training sessions.

The second hypothesis tested in this experiment, that patients have poorer kinaesthetic sense than normals, was not confirmed. Thus, while patients may still show poorer proprioceptive sense than normals in areas other than the kinaesthetic one, this experiment would seem to call such a contention into question.

As far as the Mueller-Schumann theory was concerned, the evidence was inconsistent with the notion that height of lift is a cue for judgment. Neither the normals nor the patients lifted the weights to differential heights, but they were still able to discriminate the stimuli.

Exploration of other factors relevant to the judgment situation revealed a consistent difference between normals and patients in the relative amount of time devoted to the judgment of each weight. The normals appeared to use the same amount of time for each weight, while the patients took the longest time for the middlemost weight with decreasing decision times for heavier and lighter weights.

SUMMARY

A weight judgment task was administered to 16 schizophrenic patients and 16 normals for 2 sessions. The two groups did not differ from each other in ability to discriminate between the weight stimuli, but the schizophrenics shifted significantly more than the normals in the heavy anchor condition, thus providing evidence for the theory that schizophrenics are more prone to react to concrete than to abstract stimuli, as well as for the theory that schizophrenics are less able to maintain constancy in weight judgment than normals. Measures of decision time and height of lift were also obtained and used to throw light upon the differences in shift.

REFERENCES

1. BEEBE-CENTER, J. G. The law of affective equilibrium. *Amer. J. Psychol.*, 1929, **41**, 54–69.
2. GOLDSTEIN, K. Methodological approach to the study of schizophrenic thought disorder. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Los Angeles: Univer. of Calif. Press, 1944.
3. HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 297–313.
4. HELSON, H. Perception and personality. *Air Univer. USAF Sch. of Aviation Med.*, 1953. (Rep. No. 1, Project No. 21-0202-0007.)
5. HILL, C. W. Perceptual judgment as a function of mental set, anchoring point, and method of judgment. *J. exp. Psychol.*, 1953, **46**, 325–328.
6. HUNT, W. A. Anchoring effects in judgment. *Amer. J. Psychol.*, 1941, **54**, 395–403.

7. JOHNSON, D. M. Learning function for a change in the scale of judgment. *J. exp. Psychol.*, 1949, **39**, 851–860.
8. MARTIN, LILLIE J., & MUELLER, G. E. *Zur Analyse der Unterschiedsempfindlichkeit*. Leipzig: J. A. Barth, 1899.
9. MCGARVEY, HULDA R. Anchoring effects in the absolute judgment of verbal materials. *Arch. Psychol.*, 1943, **40**, No. 281.
10. POSTMAN, L., & MILLER, G. A. Anchoring of temporal judgments. *Amer. J. Psychol.*, 1945, **58**, 43–53.
11. RADO, S. Dynamics and classification of disordered behavior. *Amer. J. Psychiat.*, 1953, **110**, 406–416.
12. SALZINGER, K. Techniques for computing shift in a scale of absolute judgment. *Psychol. Bull.*, 1956, **53**, 394–401.
13. SHERIF, M. A study of some social factors in perception. *Arch. Psychol.*, 1935, **27**, No. 187.
14. SONDER, SYLVIA L. Prediction of improvement in psychotic patients. Unpublished doctor's dissertation, Univer. of Texas, 1954.
15. TRESSELT, MARGARET E., & BECKER, M. Scales of judgment and personality correlates. *J. gen. Psychol.*, 1950, **43**, 221–230.
16. WEVER, E. G., & ZENER, K. E. The method of absolute judgment in psychophysics. *Psychol. Rev.*, 1928, **35**, 466–493.
17. WILCOXON, F. *Some rapid approximate statistical procedures*. New York: American Cyanamid Co., 1949.
18. ZUBIN, J. Psychological prognosis in early and chronic schizophrenic patients. *U S P H Progress Report M 586 C*, 1954.

Received August 9, 1956.

AN ANALYSIS OF INTERPERSON CORRELATIONS AMONG THIRTY PSYCHOTICS[1]

ROBERT J. WILLIAMS[2] AND VINCENT S. MACHI[2]

Columbia University

IN RECENT years there has been a growing interest in the development of prognostic criteria in mental disorders. For the most part, early attempts in this area were concerned with the identification of prognostic factors of a clinically observable nature (3, 7). Predicting the outcome of psychotic disorders from psychological test performance is, by comparison, a fairly recent development.

Windle (11), in a review of the literature on the prognostic use of psychological tests, has observed that a majority of studies on the prognostic value of various psychological tests report that good test performance is positively correlated with outcome. However, he also noted that a number of studies report an inverse relationship between test performance and outcome. The apparent contradictions were resolved when it was demonstrated that good performance on the part of patients with an acute (short duration) psychosis, and poor performance on the part of patients with a chronic (long duration) psychosis were both related to favorable outcome (12, 15).

A test of the prognostic value of psychological measures for recovery from psychosis must take account of those conditions which intervene between test observations and outcome. It is pertinent, for example, to consider the relationship between therapy and outcome. An evaluation of the contribution of therapy to outcome must be based on the comparison of spontaneous remission rates with remission rates following the various types of therapy. There is some agreement that special therapies produce remission rates no higher than good custodial care (9, 15). However, Zubin (13) has found that special therapies hasten improvement in some cases, although in the long run the remission rate may be the same as that following custodial care. Windle (11) concludes that, with respect to prognosis, individual factors are probably more important than type of therapy.

Another intervening condition which must be considered in evaluating prognostic indices is the time which is allowed to elapse between the initial psychological test observations and the follow-up observations. Windle and Hamwi (12) noted that a considerable amount of the confusion surrounding prognostic indices may be attributed to the neglect of some observers to report the time of follow-up, as well as the duration of illness at the time of prognostic testing. Very short-term follow-ups are unlikely to offer a clear indication of the ultimate status of the patient, and of course will fail to take into account the possibility of readmission.

An opportunity to relate a variety of psychological test performances to a sufficiently distant follow-up status was provided by the psychological test and follow-up data of the Columbia-Greystone study (6). This investigation was designed originally to measure the effects of selective partial ablation of the frontal cortex. In the course of the study a great variety of psychometric observations were made on a number of chronic psychotic patients. Some were operated and some were observed as controls. Tests were made both pre- and postoperatively. The patients are still under follow-up observation.

In the present study, factorial methods were applied to the interperson correlations among 30 chronic psychotic patients, in order (*a*) to determine if, among chronic psychotics, there exists a "psychometric type" with a high probability of recovery, and (*b*) if so, to describe this "psychometric type."

The existence of a "recovery type" has been suggested by several investigators in this area. Balinsky (1), and Lourie, Pacella, and Piotrowski (5) argue that when scores on performance tests exceed scores on verbal tests, prognosis is more favorable than when the reverse relationship exists. Pfeffer (8), Zubin and Windle (14), and Zubin, Windle, and Hamwi (15) have hypothesized that chronic patients who tend to be conceptually clear but perceptually confused have poor prognosis. These two hypotheses are not necessarily con-

[1] Portions of this article were presented at the annual American Psychological Association convention, September 3, 1956.

[2] Now with Alfred Politz Research, Inc.

tradictory and may simply be alternative expressions of the same idea. Although the number of subjects examined in the present analysis is small, and the population from which they are drawn is difficult to define, it is hoped that this study may contribute a little to the clarification of these notions.

Because the conditions under which the findings presented here could not readily be replicated, the results must be conservatively interpreted. However, an exhaustive analysis of a small number of patients may add something to knowledge in this area.

METHOD

Subjects

Thirty chronic psychotic patients of the Columbia-Greystone study were used in this analysis. Table 1 presents the age, sex, and date of first hospital admission of each patient. For the purposes of the original investigation, some of the patients were given special therapy in the form of psychosurgery. The last column of Table 1 shows whether or not each patient received this treatment. The patients are identified by the same code numbers as those used to identify the patients in the original Columbia-Greystone study. In the table, the patients are divided into two groups: those who were judged "recovered" by a criterion described below, and those who were judged "nonrecovered."

A "chronic" psychosis is defined, for the purposes of the present study, as one of two years' duration or longer. All of the present subjects (*Ss*) had been hospitalized for at least two years at the time of testing with a mean duration of 10¾ years.

Nearly all *Ss* for whom reasonably complete histories could be obtained showed evidence of prehospitalization disturbances of behavior, either in the form of poor social adjustment (extremely labile work histories, unsatisfactory marital adjustment ending in divorce or separation, or alcoholism), or in the form of abnormal episodes (reported as "convulsions," "fits," "tantrums," "depressions" or attempts at suicide). Ten of the *Ss* had records of hospitalization for behavior disturbances prior to their present admission. The available personal history data for each patient are reported by Mettler (6, pp. 315–389).

TABLE 1
DESCRIPTIVE DATA FOR 30 CHRONIC PSYCHOTIC PATIENTS OF THE COLUMBIA-GREYSTONE STUDY

Patient Code Number	Date First Hospital Admission	Age at Time of Study	Sex	Psychosurgery
"Recovered" Patients				
2	Nov., 1938	46	M	Operated
4	1934	55	M	Operated
5	Feb., 1945	35	M	Operated
7	Dec., 1942	40	M	Operated
11	Jan., 1943	44	M	Operated
12	Nov., 1940	37	M	Nonoperated
13	Oct., 1944	43	M	Operated
19	Mar., 1933	50	M	Operated
21	Nov., 1926	42	M	Operated
38	1925	38	F	Operated
39	Nov., 1924	44	F	Nonoperated
"Nonrecovered" Patients				
1	Dec., 1915	55	M	Nonoperated
3	July, 1940	61	M	Operated
6	Nov., 1940	32	M	Operated
8	Feb., 1942	29	M	Operated
9	1928	44	M	Nonoperated
14	Jan., 1933	47	M	Nonoperated
17	Nov., 1904	60	M	Nonoperated
18	July, 1943	39	M	Operated
23	June, 1942	39	M	Operated
25	Aug., 1945	58	M	Operated
26	June, 1944	35	M	Nonoperated
29	May, 1945	20	M	Nonoperated
31	June, 1941	51	F	Operated
32	June, 1934	41	F	Operated
35	Nov., 1940	59	F	Nonoperated
36	Apr., 1945	53	F	Operated
40	June, 1943	29	F	Operated
42	July, 1928	40	F	Operated
47	Jan., 1942	31	F	Operated

Psychometric Materials

Sixty-six measures of psychological test performance were made on each patient. These measures were made as a part of the Columbia-Greystone project during the Spring of 1947. Only the initial or preoperative measures were used in the present study. A summary list of the measures by name is presented in Table 2. Complete details of each test and scoring procedures are available in the literature (6, 12, 15).

Procedure

The major steps in the analysis of the data were (*a*) preparation of a matrix of interperson correlations for all thirty patients, (*b*) factorial analysis of the matrix, (*c*) identification of the factors in terms of test performances, and (*d*) determinations of the relationship of the factors to recovery and of the interrelationships of psychosurgery, age, and sex with recovery. These steps are outlined in detail below.

Preparation of the matrix. The distribution of raw scores for each of the 66 measures was dichotomized at the median. Each patient was classified as scoring either above or below the median on each measure. From these classifications, a matrix of coincident above-median scores for all possible pairs of patients was prepared. This matrix was converted into a correlation matrix by converting the coincidence scores and the number of above-median scores for each patient into percentages, and entering the tables of tetrachoric correlation coefficients (4) with these values. The use of the tetrachoric correlation coefficient involves several assumptions which cannot be verified in the present data. However, the tetrachoric coefficients were found to agree very closely with several other coefficients which were computed for portions of the data and closely duplicated the hierarchical order in the original

TABLE 2
PSYCHOLOGICAL TEST MEASURES

Test Code Number	Test Measure
1	Word Association Total Reproduction Reaction Time
2	Word Association Test Reaction Time
3	Word Association Correct Recall
4	Word Association Correct Recognition
5	Word Association Inaccurate Reproduction
6	Rorschach Total Responses
7	Rorschach Whole Responses
8	Rorschach Detail Responses
9	Rorschach Form Responses
10	Rorschach Movement Responses
11	Rorschach Color Responses
12	Rorschach Reaction Time
13	Rorschach Ratio
14	Levy Movement Cards, Scale 3 (Seeing Persons)
15	Levy Movement Cards, Scale 4 (Seeing Movement)
16	Anxiety Inventory, Full Scale
17	Anxiety Inventory, Self Descriptive Scale
18	Complaint Inventory, Full Scale
19	Social Ethical Attitudes
20	Mirror Drawing Blocking
21	Mirror Drawing Frustration
22	Complex Reaction Time Test Confusion
23	Wechsler Bellevue General Information
24	Wechsler Bellevue General Comprehension
25	Wechsler Bellevue Digit Span
27	Wechsler Bellevue Similarities
28	Wechsler Bellevue Vocabulary
29	Wechsler Bellevue Picture Arrangement
30	Wechsler Bellevue Picture Completion
31	Wechsler Bellevue Block Design
32	Wechsler Bellevue Object Assembly
33	Wechsler Bellevue Digit Symbol Substitution
34	Wechsler Bellevue Full IQ
35	Porteous IQ
36	Porteous MA
37	Porteous Q Score
38	Complex Reaction Time Test, Number of Correct Solutions
39	Complex Reaction Time Test, Number of Errors
40	Albert Test, Number Correct
41	Analogies Test, Number Correct
42	Essential Differences Test, Number Correct
43	PI Sorting Test, Part 1, Number Grouped
44	PI Sorting Test, Part 2, Verbalized Groups
45	PI Sorting Test, Part 3, Isolation of Single Items
46	Weigl Test Performance
47	Capps Homograph Test, Verbal Flexibility Score
48	Benton Visual Retention Test, Number Correct
49	PI Sorting Test, Part 4, Recall
50	Memory For Objects, Recall
52	Memory For Objects, Recognition
53	Subtraction Time
54	Addition Time
55	Critical Flicker Frequency
56	Time Judgment, Total Discrepancy
57	Time Interval Test, Reproduction Discrepancy
58	Time Interval Test, Verbal Estimation Discrepancy
59	Time Interval Test, Operative Estimation Discrepancy
60	Time Interval Test, Discrepancy sans Sound
61	Time Interval Test, Discrepancy with Sound (Buzzer)
62	Mirror Drawing, Total Sides
63	Mirror Drawing, Error
64	Meaningful Paired Associates, Number of Learning Trials
65	Semi-meaningful Paired Associates, Number of Learning Trials
66	Meaningful Paired Associates, Variability in Learning
67	Semimeaningful Paired Associates, Variability in Learning
68	Verbal Directions, Rate of Learning

coincidence scores. Ease of computation therefore favored the use of the tetrachoric coefficients.[3]

The factor analysis. The matrix of interpatient correlations was factored, using Thurstone's complete centroid method with the highest correlation in each column or row entered as an estimate of the communality for each patient. Factoring was terminated when the matrices of residuals met Tucker's criterion (2, p. 298).

At this point, criterion data were obtained for all patients. The criterion is based on the hospital status of patients as of May 5, 1954. On this date, of the 30 patients considered in this study, 17 were hospitalized, 8 were discharged, 4 had died, and one patient had been deported. Of the four who died, one died while hospitalized, and three died after discharge. The patient who died in the hospital died on May 22, 1952 after 19 years of hospitalization. This patient was assigned a hospital status of "Hospitalized." Of the three patients who died after hospital discharge, one (patient 4) died on April 9, 1952 after 4½ years of nonhospitalization, another (patient 13) died January 12, 1954 after 6½ years of nonhospitalization, and the third (patient 19) died December 4, 1949 fifteen months after leaving the hospital. Each of these three patients was assigned a hospital status of "discharged." The deported patient (patient 9) was assigned a hospital status of "hospitalized" on the basis of his deportation from Greystone in July, 1951, because of "mental illness." Thus, with a criterion of hospital status as of May 5, 1954, there was a total of 11 "discharged" or "recovered" patients, and a total of 19 "hospitalized" or "nonrecovered" patients. In Table 1, patients are classified as "recovered" and "nonrecovered" by this criterion.

An orthogonal rotation of the original factors was then carried out so as to maximize the projection of the centroid of the "recovered" group on Factor I, while minimizing its projection on all other factors.

Identification of the factors. Patients were selected to represent the extreme positive and extreme negative ends of each factor. Care was taken to select these patients so that their average loading on factors other than the one they were selected to represent would be a minimum. Above-median performance was assigned an arbitrary value of +1, and below-median performance an arbitrary value of −1. The performance of each selected patient on each of the 66 measures was then weighted by his loading on the factor which the patient was chosen to represent, and the products were summed over patients for each measure. These summed weights were used to rank the measures as contributors to each factor, so that the factors could then be described in terms of the tests. The factor of principal interest in the present study is the one which was rotated to pass through the centroid of the "recovered" group. The other factors were described only for the purpose of clarifying the description of this "recovery factor." These additional descriptions serve the purpose of showing what the "recovery factor" is *not*.

Interrelations of criterion data. The rotation of one of the factors to a criterion of maximizing the contrast between recovered and nonrecovered patients tends to capitalize on chance factors which may contribute to

[3] Since this analysis was completed, Dr. Joseph Zubin has suggested that the coincident above-median scores, corrected for chance, might have been factored directly.

TABLE 3
INITIAL AND ROTATED FACTOR LOADINGS OF THIRTY CHRONIC PSYCHOTIC PATIENTS

	UNROTATED				ROTATED				
	I	II	III	IV	I	II	III	IV	h^2
Recovered									
2	−.340	−.093	.246	.731	.360	.289	.271	.658	.720
4	.508	.513	.044	.293	.683	−.308	−.043	−.213	.609
5	.581	.184	.052	.289	.469	−.486	.002	−.029	.458
7	.383	−.264	−.046	−.160	−.200	−.449	−.029	−.036	.244
11	−.046	.382	.215	.052	.312	.171	.167	−.207	.197
12	.176	−.312	.464	.520	.216	−.270	.491	.503	.614
13	.638	−.238	−.134	.169	.090	−.681	−.131	.146	.510
19	−.230	.582	.150	−.254	.197	.411	.086	−.513	.478
21	−.322	.437	.191	−.243	.084	.449	.149	−.399	.390
38	−.180	.051	−.756	.151	.016	.187	−.748	.188	.630
39	−.591	.375	−.190	−.020	.084	.682	−.209	−.101	.526
Nonrecovered									
1	−.556	−.444	.401	−.064	−.451	.376	.478	.313	.671
3	−.506	−.262	−.253	−.437	−.623	.390	−.195	−.032	.579
6	.408	−.195	−.435	−.201	−.212	−.450	−.425	−.081	.435
8	.189	.144	.414	.145	.285	−.130	.384	−.059	.249
9	−.321	−.561	−.112	.138	−.392	.115	−.026	.531	.450
14	−.302	−.402	−.241	−.263	−.554	.151	−.175	.139	.380
17	−.422	.464	.088	.355	.464	.553	.048	.063	.527
18	−.462	.044	−.222	.348	.125	.450	−.205	.354	.385
23	−.394	−.240	.038	−.134	−.350	.291	.085	.131	.232
25	.720	.433	.307	.176	.628	−.534	.218	−.321	.830
26	.407	−.158	−.452	.172	.059	−.436	−.446	.177	.424
29	.400	−.127	.211	−.382	−.222	−.419	.207	−.314	.366
31	.223	.215	−.281	−.261	.004	−.138	−.315	−.353	.243
32	−.558	−.189	−.027	−.475	−.588	.463	.022	−.114	.574
35	−.289	−.142	.041	.056	−.130	.225	.071	.190	.109
36	.309	.702	.049	−.217	.429	−.057	−.053	−.669	.638
40	.383	−.320	.197	−.370	−.355	−.468	.218	−.180	.425
42	−.047	.341	−.225	.113	.279	.158	−.264	−.096	.182
47	.654	−.162	−.267	.141	.115	−.671	−.273	.086	.545

that contrast. Therefore, no genuine test of the significance of the relationship between this factor and "recovery" is possible. However, White's rank order test (10) was applied with this serious reservation.

The possibility that the "recovery factor" might be an artifact of age or sex differences was examined by determining the relation of these variables to the factor loadings, using White's rank test for sex and the rank-difference correlation for age. (An examination of the descriptive data in Table 1 shows that the "recovery" criterion used in this study is unrelated to duration of illness, age, sex, or psychosurgery.)

RESULTS[4]

Four factors were extracted from the matrix of interperson correlations. The matrix of residuals after the extraction of these four factors did not warrant the extraction of additional factors according to Tucker's criterion. These factors were then rotated as described above so that the first factor would pass through the centroid of the "recovered" group. The original and rotated factors are presented in Table 3, where "recovered" and "nonrecovered" patients are separately grouped. With but one exception (patient 7), all of the "recovered" patients have positive loadings on the rotated Factor I. Of the 19 "nonrecovered" patients, 10 have negative loadings on Factor I, and 9 have positive loadings. The difference between Factor I loadings of the "recovered" and "nonrecovered" patients is significant at the .05 level by White's rank-order test ($T = 122$ $n_1 = 11$, $n_2 = 19$). (This finding must be conservatively interpreted because of the reservations noted earlier concerning the application of any test of significance to these data.)

[4] The following tables, too bulky to be reproduced here, have been deposited with the American Documentation Institute: (*a*) The above- or below-median classification of each patient on each of the 66 measures, (*b*) the matrix of coincident above-median scores, (*c*) the matrix of initial intercorrelations, (*d*) the matrix of residuals, (*e*) the transformation matrix for rotating the factors, and (*f*) a table showing the rank, for each factor, of each of the 66 measures. Order document No. 5197 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm and $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

No relationship was found between recovery–nonrecovery and psychosurgery. A chi-square test of the independence of these two principles of classification for these patients yielded a value of 2.302 with one degree of freedom, P lying between .20 and .10.

Factor I was found to be unrelated to age (rho = −.11, not significantly different from zero). This factor was also found to be unrelated to sex by White's rank test ($T' = 125.5$, $n_1 = 9$, $n_2 = 21$, not significant). Similarly, the criterion of "recovery" is unrelated either to age (a chi-square test of the independence of the criterion and age categories of 20–34 years, 35–49 years, and 50–64 years, yielded a value of 1.176 with 2 *df*, P falling between .7 and .5), or sex (chi-square of .437 with 1 *df*, P between .7 and .5). Neither the criterion nor Factor I is related to duration of illness at time of testing. No relation was anticipated since these patients were selected as having chronic (or long duration) psychoses.

Table 4 lists the 20 measures which contributed most to a definition of each factor. The measures are listed in a decreasing order of importance (as defined above in the Procedure section) as contributors to the factor. The minus signs in front of some test code numbers

TABLE 4

PSYCHOLOGICAL MEASURES WHICH CONTRIBUTE MOST TO A DEFINITION OF EACH FACTOR

FACTOR I	FACTOR II
−42 Essential Differences, Correct	−5 Word Association, Incorrect
−48 Benton Visual Retention, Correct	−53 Subtraction Time
−31 WB Block Design	23 WB Information
−24 WB Comprehension	28 WB Vocabulary
54 Addition Time	47 Capps Homographs
−30 WB Picture Completion	49 PI Sorting, Recall
−27 WB Similarities	68 Verbal Directions, Rate of Learning
−38 Complex Reaction Time, Correct	−19 Social-Ethical
−34 WB Full IQ	45 PI Sorting, Single items
37 Porteous Q score	3 Word Association, Correct
−23 WB Information	−8 Rorschach, Detail
22 Complex Reaction Time, Confusion	−10 Rorschach, Movement
−29 WB Picture Arrangement	−1 Word Association, Reaction Time
−59 Time Interval, Operative Est.	46 Weigl Performance
−60 Time Interval, Silent	−58 Time Interval, Verbal
−61 Time Interval, Sound	−59 Time Interval, Operative
−25 WB Digit Span	−60 Time Interval, Silent
−28 WB Vocabulary	−64 Meaningful Paired Associates, Trials to learn
19 Social-Ethical	−65 Semimeaningful Paired Associates, Trials to learn
53 Subtraction Time	27 WB Similarities

FACTOR III	FACTOR IV
−15 Levy Movement Cards	−21 Mirrow Drawing, Frustration
6 Rorschach, Total Responses	−67 Paired Associates, Semimeaningful Variability
20 Mirrow Drawing, Blocking	−16 Anxiety A
23 WB Information	−17 Anxiety B
29 WB Picture Arrangement	−22 Complex Reaction Time, Confusion
53 Subtraction Time	−37 Porteous Q Score
65 Semi-Meaningful Paired Associates, Trials to learn	−47 Capps Homographs
−2 Word Association, Reaction Time	−68 Verbal Directions, Rate of Learning
−19 Social-Ethical	29 WB Picture Arrangement
−37 Porteous Q	40 Albert Test
−39 Complex Reaction Time, Errors	20 Mirror Drawing, Blocking
43 PI Sorting, Number Grouped	−50 Memory for objects, Recall
48 Benton Visual Retention, Correct	−39 Complex Reaction Time, Errors
57 Time Interval Reproduction	9 Rorschach, Form
62 Mirror Drawing, Total sides	34 WB Full IQ
−22 Complex Reaction Time, Confusion	2 Word Association, Reaction Time
8 Rorschach Detail	4 Word Association, Correct
63 Mirror Drawing, Error	33 WB Digit Symbol
−44 PI Sorting, Verbalized	31 WB Block Design
−45 PI Sorting, Single Items	−52 Memory for Objects, Recognition

indicate that *below*-median performance on those tests contributed to a definition of the positive end of that factor. For tests without minus signs, *above*-median performance contributed to a definition of the positive end of that factor.

The relation of Factor I to the recovery criterion shows that patients who eventually "recover" almost always have high loadings on this factor, although patients with high loadings on this factor do not invariably recover. For this reason an additional comparison was carried out. The performance of the five "recovered" patients with highest loadings on Factor I (patients 2, 4, 5, 11 and 12) was compared with the performance of five "nonrecovered" patients with highest loadings on that factor (patients 8, 17, 25, 36 and 42), using a procedure similar to that which was applied in obtaining descriptions of the factors in terms of the tests, except that the performances were not weighted by the factor loadings. It was found that the recovered patients more often than the nonrecovered show superior performance on Seeing Persons and Seeing Movement in the Levy movement cards, the Albert test, Rorschach Movement, the Picture Arrangement, Object Assembly, and Digit-Symbol Substitution subtests of the Wechsler-Bellevue, and the PI Sorting test. The recovered patients show inferior performance on Word Association Reaction Time, Time Interval Reproduction, and Rorschach Reaction Time. Nonrecovered patients, on the other hand, are superior to the recovered patients on Verbal Directions, WB Vocabulary, Porteous IQ and MA, the Analogies Test, the Weigl test, the Capps Homograph test, and Verbalized Groupings on the PI Sorting test. The nonrecovered group required fewer trials to learn Semimeaningful Paired Associates, but had greater "confusion" scores on Complex Reaction Time.

DISCUSSION

Factor I is associated with all of the intelligence tests. In its positive or "recovery" aspect, it is associated with poor performance on these tests as well as with long reaction times and poor time judgment. Factors II, III, and IV are also associated with "intelligence" measures, and the components of intelligence involved in these factors are more easily identified than in Factor I. Therefore, a practical approach to a description of Factor I may be first to describe what it is *not* in terms of Factors II, III, and IV.

Factor II in its positive aspects appears to represent good memory and concept formation. It is clearly a verbal factor as indicated by the high relative importance of the General Information, Vocabulary, and Similarities subtests of the Wechsler-Bellevue in its positive aspects, and Word Association Inaccurate Reproduction on its negative side. It also represents concept formation at the verbal level as indicated by the high relative importance of the PI Sorting-Recall and Isolation of Single Items, and Weigl Test Performance. Some verbal flexibility is indicated by the Capps Homograph Score. Factor II may be thus described as containing various verbal components of intelligence.

Factors III and IV both appear to involve perceptual abilities. Factor III is associated with Rorschach Total Responses and Detail, Mirror Drawing scores (scores 20, 62, and 63), Picture Arrangement subtest of the Wechsler-Bellevue, the Benton Visual Retention Test, and Seeing Movement in the Levy Movement cards. Patients who have high loadings on this factor seem capable of good performance on the PI Sorting tests, but are unable to verbalize their groupings. Factor IV is associated with the Picture Arrangement, Block Design, and Digit Symbol subtests of the Wechsler-Bellevue and Rorschach Form responses. Certain verbal tests contribute negatively to this factor, such as the Capps Homograph test and the Wechsler-Bellevue Similarities and Vocabularly. Also contributing negatively are Mirror Drawing Frustration and both Recall and Recognition Memory for Objects. Both Factors III and IV, then, seem to contrast good "perceptual" performances with poor "verbal" performances. The first seems to deal with the perception of concrete and familiar objects, whereas the latter is concerned with the perception of somewhat more abstract designs.

Factor I, then, cannot represent the verbal and perceptual components of test performance. The positive end of Factor I, which is associated with "recovery," represents low scores on all of the various tests of "intelligence." However, there is reason to believe that poor performance results from the failure of these patients to respond to the test situa-

tion. Confusion and sluggishness is indicated by the incorrect estimation of time intervals, and the various long reaction time scores. Poor test performance, probably brought about by failure to respond to the test situation, seems then to be a necessary condition for recovery from chronic psychosis.

Among the patients who had high loadings on Factor I, the "recovered" group is distinguished from the "nonrecovered" group in that the "recovered" patients seem better able to perform well on "perceptual" tasks (Levy Movement cards, Albert test, Rorschach Movement, WB Picture Arrangement, Object Assembly and Digit Symbol, and the PI Sorting test) which require some learning in the immediate situation. The nonrecovered patients are apparently better able to make use of previous learning, particularly of a verbal-conceptual nature (Verbal Directions, Vocabulary, Porteous IQ and MA, Analogies, and Homographs). These findings lend some support to the hypothesis that prognosis is favorable when "perceptual" scores exceed "conceptual" scores (8, 14, 15). However, this hypothesis concerns intrapatient comparisons and the above findings are based on interpatient comparisons. Also, because they are based on comparisons among only ten patients, they must be interpreted with some caution.

Thus, the conditions favorable for a prognosis of recovery from chronic psychosis appear to include (*a*) generally poor performance on tests of intellectual abilities, (*b*) confusion and sluggishness as evidenced in long reaction times and poor time estimation, and (*c*) somewhat better performance in perceptual tasks requiring learning in the immediate situation over performance on verbal-conceptual tasks which depend on prior learning.

SUMMARY

The purpose of this study was (*a*) to determine if, among chronic psychotics, there exists a "psychometric type" with a higher than chance probability of recovery, and (*b*) if so, to describe that type.

The subjects were 30 chronic psychotics of the Columbia-Greystone study, 11 of whom were classified as "recovered" as of May 5, 1954.

A matrix of the interperson correlations among all subjects based on above-median scores on 66 psychological test measures was factor analysed by Thurstone's complete centroid method. Four factors were extracted. These were rotated to maximize the projections of the "recovered" group on one factor while minimizing their projections on all other factors. The conclusions drawn from the analysis are as follows:

1. Among chronic psychotics there exists a "psychometric type" with a better-than-chance probability of recovery. This type is not distinguished by age, sex, or duration of illness.

2. The type is characterized by generally poor performance on intellectual tasks, long reaction times, and poor time estimation.

3. Among patients of this type, those who recover perform better than those who do not on perceptual tasks requiring new learning, whereas those who do not recover perform better than those who do on verbal-conceptual tasks depending on previous learning.

REFERENCES

1. BALINSKY, B. Factors in the vocational adjustment of schizophrenics after mental hospital discharge. *J. clin. Psychol.*, 1947, **3**, 341–349.
2. CATTELL, R. B. *Factor analysis; an introduction and manual for the psychologist and social scientist.* New York: Harper, 1952.
3. CHASE, L. S., & SILVERMAN, S. Prognostic criteria in schizophrenia; a critical survey of the literature. *Amer. J. Psychiat.*, 1941, **98**, 360–368.
4. CHESIRE, L. S., SAFFIR, M., & THURSTONE, L. L. *Computing diagrams for the tetrachoric correlation coefficient.* Chicago: The Univer. of Chicago Bookstore, 1933.
5. LOURIE, R. S., PACELLA, B. L., & PIOTROWSKI, Z. A. Studies on the prognosis in schizophrenic-like psychoses in children. *Amer. J. Psychiat.*, 1943, **99**, 542–551.
6. METTLER, F. A. (Ed.) *Selective partial ablation of the frontal cortex.* New York: Hoeber, 1949.
7. PASCAL, G. R., SWENSEN, C. H., FELDMAN, D. A., COLE, M. E., & BAYARD, J. Prognostic criteria in the case histories of hospitalized patients. *J. consult. Psychol.*, 1953, **17**, 163–171.
8. PFEFFER, M. Pattern of abilities in chronic schizophrenics. Unpublished doctor's dissertation, Columbia Univer., 1952.
9. SCHERER, I. W. Prognoses and psychological scores in electroconvulsive therapy, psychosurgery, and spontaneous remission. *Amer. J. Psychiat.*, 1951, **107**, 926–931.
10. WHITE, C. The use of ranks in a test of significance for comparing two treatments. *Biometrics*, 1950, **8**, 33–41.

11. WINDLE, C. Psychological tests in psychopathological prognosis. *Psychol. Bull.*, 1952, **49**, 451–482.
12. WINDLE, C., & HAMWI, V. An exploratory study of the prognostic value of the Complex Reaction Time Test in acute and chronic psychotics. *J. clin. Psychol.*, 1953, **9**, 156–161.
13. ZUBIN, J. Evaluation of therapeutic outcome in mental disorders. *J. nerv. ment. Dis.*, 1953, **117**, 95–111.
14. ZUBIN, J., & WINDLE C. Psychological prognosis of outcome in the mental disorders. *J. abnorm. soc. Psychol.*, 1954, **49**, 272–281.
15. ZUBIN, J., WINDLE, C., & HAMWI, V. Retrospective evaluation of psychological tests as prognostic instruments in mental disorders. *J. Pers.*, 1953, **21**, 342–355.

Received August 13, 1956.

FREQUENCY AND IDIOSYNCRACY OF ASSOCIATIVE RESPONSES[1]

GEORGE MANDLER AND ELIZABETH W. PARNES

Harvard University

INDIVIDUAL differences in response to environmental stimuli can be described in terms of the frequency and the normative aspects of the responses elicited by those stimuli. Thus, the response to or perception of a stimulus is determined by the particular array of responses which that stimulus elicits from the subject (*S*). The responses which thus differentiate the stimulus from other stimuli represent not only its immediate "meaning," but can be conceptualized as the result of the *S*'s previous learning experiences with such stimuli (4). It has also been argued that this acquisition of differentiating responses forms the basic paradigm for perceptual learning (5).

Empirically, it has been shown that the frequency of responses elicited by verbal stimuli is related to familiarity with the stimulus (10) and to speed of serial and paired-associate learning (2, 7, 9). Noble (8) and Mandler (4, 6) have suggested a relationship between the associative frequency of a stimulus and its meaningfulness, theoretically defined or empirically determined. Finally, it has been shown that individual rates of learning are related to the subject's frequency of emitted responses (7).

This paper reports two studies of the consistency of individual differences in frequency of response to different stimuli and the effect of stimulus variation upon response emission. Five general hypotheses are examined:

1. There are significant differences among *S*s in both frequency and idiosyncratic content of associative responses.

2. The *S*s show consistent tendencies in the number of responses evoked by a variety of different stimuli.

3. There are consistent tendencies in the content of the responses evoked by different stimuli, defined as the relative frequency of idiosyncratic or individual-specific responses.

4. Stimulus classes differ from each other in the elicitation of responses in respect to both frequency and to content.

5. A failure-stress situation produces measurable differences in individual response output and idiosyncratic content.

It is patently impossible to elicit and to quantify free association material in a controlled laboratory situation. However, this study represents an initial attempt to evoke controlled associations to a variety of different materials. The term "controlled" rather than "free" is used since the *S* is instructed to give those responses which are *directly* elicited by the stimulus. In this manner, a first approximation to individual differences in associative material may be obtained. The materials used in this preliminary study were selected to represent a diversified range of visually presented materials. Both verbal and nonverbal stimuli were included in order to ascertain any effect which might be directly related to response differences directly attributable to the verbal nature of the stimulus materials. In addition, it was decided to include some stimuli which are part of the clinical armamentarium in order to ascertain any differences between fairly simple stimulus materials and those classically believed to represent the best conditions for the elicitation of personally meaningful material.

In line with these considerations, four classes of stimulus materials were selected. The two types of verbal stimuli consisted of nonsense syllables and adjectives. It has experimentally been shown that differences in the associative frequency of nonsense syllables are empirically useful, while adjectives were selected as representing verbal materials more closely related to the *S*'s everyday learning experiences. The nonverbal stimuli consisted of line drawings and Rorschach cards. The line drawings, ranging from simple objects to nonsense figures, again tap a more familiar array of stimuli, while the Rorschach cards represent a clinically useful and relatively unstructured stimulus situation.

In order to assess further the usefulness of this technique in assessing personality variables, it was decided to introduce a simple yet widely used experimental procedure. Failure-stress was selected for this part of the study

[1] This study was supported by grants from the Laboratory of Social Relations, Harvard University, and from the Foundations Fund for Research in Psychiatry.

because it has been shown to be useful in the elicitation of personality differences (3).

METHOD

The first of two studies (Study I) investigated the responses of 20 *Ss* to 40 different stimuli, ten in each of four stimulus classes: Nonsense Syllables, Adjectives, Drawings, and Rorschach cards. In Study II, two groups of 12 *Ss* each were used. The control group responded to 24 stimuli (six from each of the four stimulus classes); the experimental group responded to 12 stimuli, were then given failure-stress instructions, then responded to the remaining 12 stimuli.

Materials

In Study I, the stimuli were selected from each of the four stimulus classes in order to obtain a maximum range of differentiation in meaningfulness and structure on the basis of available information.[2]

Nonsense syllables were selected from the list of 100 published by Mandler (6). This standardization provides an index of associative frequency which describes the mean number of associations elicited during a 30-sec. period. These 100 syllables were divided into ten groups of 10 syllables each on the basis of an index of ambiguity. This index is derived from the difference between the mean number of *different* responses elicited by a stimulus and the mean number of total responses elicited. Ambiguity increases as this index approaches zero. An index of zero would be obtained if no response to a particular stimulus is given by more than one *S*. One syllable was chosen at random from each of the ten ambiguity groups.

Nine adjectives were selected from each of nine different frequency categories from the Thorndike-Lorge word list. They ranged from more than 100 per one million words to 4 per 18 million words. The tenth adjective was selected from an unabridged dictionary as one appearing both unusual and unfamiliar to the experimenters.

All 10 drawings were selected on the basis of unpublished data collected by Baum[3] from the pictures used by Heidbreder in her concept formation studies (1). The range of drawings was from simple realistic objects to abstract complex figures, and selection was based on the number of different single responses elicited by the stimuli on two subsequent trials. The selected drawings ranged from 11 per cent to 86 per cent in the percentage of different responses elicited from a group of *Ss*, and from 11 per cent to 73 per cent in the percentage of *Ss* who changed the response from trial one to trial two, thus covering a wide range of response variability.

The ten standard Rorschach cards were also used.

In Study II, only six stimuli were used in each of the four stimulus classes. The selection procedure was identical for the four classes. The stimuli within each class were ranked on the basis of the number of responses elicited in Study I, and ranks Nos. 1, 2, 5, 6, 9, and 10 were selected. In the case of tied ranks, the stimulus with the greatest number of popular responses (defined below) was chosen. The range obtained in respect to ambiguity and frequency was highly similar to that in Study I for the Syllables, Adjectives, and Drawings. This procedure resulted in the selection of Rorschach cards Nos. 3, 5, 7, 1, 6, and 8 in the order of ranks chosen.

Subjects

The *Ss* in Study I were 19 male and one female undergraduate students recruited through the University employment bureau. In Study II, the *Ss* were 12 male and 12 female summer school students. The control and experimental groups consisted of six male and six female *Ss* each, assigned at random.

Procedure

Instructions. The *Ss* in both studies were initially informed that they were to take part in an experiment designed to ascertain "what kind of responses people give to certain stimuli." They were instructed to respond to each stimulus with all the "words that the stimulus brings to mind." *Ss* were asked not to give chain-associations, but rather to be sure to give responses directly elicited by the stimulus. They were also requested not to give phrases or sentences, but to restrict themselves to single words.

In Study II, the control group proceeded after the initial instruction without any further interruptions. The experimental group was interrupted after the 12th stimulus presentation. At that time the experimenter (*E*) informed them that they were being interrupted because they were doing so poorly and that they either showed low ability or were "not trying at all." They were asked to buckle down on the rest of the stimuli. No other information was given them. They then proceeded with the second half of the stimulus presentation.

Stimulus presentation and assignment. The stimulus materials were presented by means of a modified juke box mechanism, which allowed the *E* to select the serial presentation in a predetermined order. Stimuli were mounted on 10-in. discs and presented to the *S* through a mirror arrangement. All stimuli were presented for 30 sec. with an interval between stimuli of approximately 10 sec.

In Study I, each *S* received a different order of stimulus presentations. The order of the 40 stimuli was random except that each block of four successive stimuli contained one item from each of the four stimulus classes, and no two items of the same stimulus class followed each other.

In Study II, twelve such series were prepared. Thus, the 12 control *Ss* and the 12 *Ss* in the experimental group received identical serial presentations, with the additional provision that *Ss* receiving the same stimulus order were of the same sex.

[2] The following materials have been deposited with the American Documentation Institute: Lists of stimulus materials used; the initial correlation matrix for the factor analysis; tables of response frequency and idiosyncratic percentage for the four stimulus classes for the 20 *Ss* in Study I. Order Document No. 5198 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting $1.25 for 35 mm. microfilm or $1.25 for 6 x 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[3] Thanks are due Dr. Marian H. Baum, of Oberlin, Ohio, for making these data available.

Dependent Variables

The responses of all *S*s were recorded and transcribed. For purposes of analysis, a response was defined in the following way:

1. Comments to the *E*, exclamations, connectives, and other obviously irrelevant responses were excluded.

2. When the *S* responded with a phrase (contrary to the instructions), the response was categorized by the main noun, if any; otherwise, by the main verb. The main word was defined as appearing in the major clause of a sentence.

3. For counting replications of a response by *S* on a particular stimulus, a response was tallied only once if the replications represented plural or singular forms of the same noun, different forms of the same verb, or verbs and adjectives with identical stems. Thus, a single response is defined as a single word, complex word, or phrase. For coding purposes, responses were reduced to single words.

These responses were tabulated separately for each stimulus. In order to establish frequency distributions of responses given to a particular stimulus by all *S*s, responses were combined, i.e., counted as replications, in a manner similar to that used for replications by *S*s.

In Study I, the following measures were obtained from these tabulations for each *S*:

F—the number of responses given.

I—the number of idiosyncratic responses given, defined as those responses which are given to a particular stimulus by one, and only one, *S*.

I%—the percentage of idiosyncratic responses out of the total number of responses.

P—the number of popular responses given, defined as those responses which are given to a particular stimulus by 40 per cent or more of the *S*s.

P%—the percentage of popular responses out of the total number of responses.

P′—the number of popular responses given as initial responses to the stimuli.

T—total popularity measure of responses. T for any one *S* is defined as the sum of the index of popularity of all responses. This index is the number of *S*s giving that response. Thus, the T value for an idiosyncratic response is one, of a response given (to a particular stimulus) by 12 *S*s is 12.

T/F—mean popularity index.

X%—percentage of responses which are neither popular nor idiosyncratic.

A—index of popularity (see T) of all *initial* responses given by *S*.

M—the number of specific responses which are given by *S* to more than one stimulus.

M′—M times the number of times it is given by *S*.

M′%—the percentage of M′ responses.

In Study II, only the F and I% measures were used.

RESULTS

The analysis of the data proceeded in three general steps. The first consisted of a factor analysis of the response measures listed in the Method section. On the basis of this analysis, the two variables of response frequency and percentage of idiosyncratic responses were selected for more detailed testing of the five experimental hypotheses.

Factor Analysis[4]

The 13 variables listed above were used to obtain a correlation matrix based on the scores of the 20 *S*s used in Study I.[5] For all measures, the scores computed for each *S* were obtained from all the 40 stimuli used for all *S*s. Two factors were obtained, and one orthogonal rotation resulted in the factor loadings shown in Table 1.

The two factors can be relatively easily identified. Factor A shows the highest loadings for number of popular responses (P), number of popular responses given initially (P′), popularity of all initial responses (A), the mean index of popularity of all responses (T/F), and a high negative loading in the percentage of idiosyncratic responses (I%). This cluster of variables leads to the designation of the first factor as a popularity dimension, i.e., the tendency to give common responses or to give a low percentage of idiosyncratic responses. Factor B shows the highest loadings in the frequency of response (F), total number of idiosyncratic responses (I), and the tendency to repeat identical responses to different stimuli (M and M′). All of these variables can be conceptually related to *S*s' general frequency of response output, and the factor can be designated as a frequency dimension. On the basis of these two factors, one variable was selected from each of the two for further analysis. Frequency of response (F) was selected as representative of Factor B and percentage of idiosyncratic response (I%) as representative of Factor A. The selection of I% was based not only on the high negative loading this variable showed on Factor A, but also because it provided a more convenient scoring system for Study II than the other popularity measures.

Response Frequency

The individual difference variable showed a high degree of subject-to-subject variability in both studies. In Study I, the mean number of

[4] Dr. Ray Hyman's assistance in this analysis is gratefully acknowledged.

[5] See Footnote 2.

TABLE 1
ROTATED FACTOR LOADINGS FOR ALL VARIABLES IN STUDY I

Variable	Factor Loadings	
	Factor A	Factor B
F	+.31	+.89
I	−.11	+.91
I%	−.75	+.44
P	+.86	+.18
P%	+.60	−.66
P′	+.86	+.20
T	+.66	+.65
T/F	+.78	−.55
X%	+.70	−.07
A	+.86	+.12
M	+.14	+.96
M′	+.13	+.97
M′%	−.05	+.69

responses elicited by a stimulus in a 30-sec. period was 6.9.[6] Subject means ranged from 2.6 to 16.0 with a standard deviation of 3.63. In Study II, the mean was 6.0, range 2.5 to 15.1, standard deviation 3.52. Thus, this variable provides a sensitive differentiation between individuals, with the most responsive *S*s producing as much as six times as many responses on the average than the least responsive *S*s.

Analyses of variance, summarized in Table 2, were performed on the frequency data from both studies. In both studies, the largest contribution to variance was from differences between *S*s. Stimulus classes and stimulus differences within classes (in Study I), however, were highly significant sources of variation. Additional analysis of Study I revealed that the four stimulus classes chosen showed greater variability between classes than within classes ($F = 6.47$, $p < .005$). Similarly, differences between *S*s obviously contributed more to variance than differences between the stimuli used. Finally, the interaction between individuals and stimulus classes in Study I was significantly greater than the interaction between individuals and stimuli within classes. This finding indicates that not only do *S*s react differently to different kinds of stimuli, but that this differentiation is greater than the differential reaction to stimuli within a particular class.

The absolute differences in frequency of responses elicited by different stimulus classes were not, however, very great. Table 3 shows the means of the various classes, and the ranges within these classes in Study I. While the means show relatively small differences between stimulus classes, the full range for the stimuli from 4.8 (for the nonsense syllable "Duj") to 8.8 (for the adjective "Cordial" and the drawing of a house) suggests substantial differences between stimuli in their power to elicit a greater or lesser frequency of response. A surprising finding was the relatively small range of differentiation of Rorschach cards when compared with the other stimulus classes.

In Study II, where stress was introduced, all *S*s' responses to stimuli within a particular class were used as the unit of response. In addition to the major effects found already in Study I, a significant source of variance was associated with the first and second halves of the experiment. This effect must be viewed in the light of the significant Halves × Conditions interaction. It will be recalled that all *S*s were presented with the same stimulus conditions controlled for serial presentation. The experimental group, however, was subjected to failure stress after the first half of the stimuli had been presented. Thus, any effect due to the introduction of this variable should show up in the Halves × Conditions interaction. This significant source of variance can be attributed to a significantly greater increase in response frequency in the experimental group. While the control group showed a mean increase of 13 per cent, the experimental group showed an increase of 30 per cent. One of the effects of stress, therefore, was to increase the *S*'s rate of responding well beyond a general increase in responsivity from the first to the second half of the experimental situation. The only other significant finding was the quadruple interaction between Stimulus Classes, Sex of *S*s, Conditions, and Halves ($p < .05$). Without prior predictions it is difficult to interpret such a complex interaction effect.

The final question to be raised concerning the frequency variable concerns the consistency of response frequency across different stimulus classes. Table 4 shows a measure, the index of concordance (W), which indicates this consistency tendency. Ws were computed for all *S*s in Study I, for the control group in Study II, and separately for the pre- and post-experimental stimuli in the experimental

[6] See Footnote 2.

TABLE 2
SUMMARY OF ANALYSES OF VARIANCE OF RESPONSE FREQUENCY IN STUDIES I AND II

Source	Study I				Study II			
	df	Mean Square	F	*p*	*df*	Mean Square	F	*p*
Total	799	17.1	—	—	191	122.0	—	—
Between *Ss*	19	533.1	48.86	<.001	20	789.3	22.41	<.001
Stimulus Classes	3	63.2	5.79	<.005	3	300.3	18.37	<.001
Stimuli within Classes	36	9.8	3.69	<.001	—	—	—	—
Ss × Stimulus Classes	57	10.9	4.11	<.001	—	—	—	—
Halves*	—	—	—	—	1	740.0	21.39	<.001
Halves × Conditions†	—	—	—	—	1	153.0	4.42	<.05

* "Halves" refers to the first and second half of Study II.
† "Conditions" refers to the experimental vs. control groups in Study II.

TABLE 3
MEAN AND RANGE OF RESPONSE FREQUENCY (F) ELICITED BY FOUR STIMULUS CLASSES

Stimulus Class	Mean	Range
Nonsense syllables	6.2	4.8–7.3
Adjectives	7.1	6.1–8.8
Drawings	7.5	6.5–8.8
Rorschach cards	6.7	6.4–7.4

TABLE 4
CONSISTENCY OF INDIVIDUAL DIFFERENCES ACROSS FOUR STIMULUS CLASSES

Source of Data	F*		I%†	
	W	*p*	*W*	*p*
Study I (20 *Ss*, 40 stimuli)	.900	<.01	.672	<.01
Study II				
Control (12 *Ss*, 24 stimuli)	.872	<.01	.691	<.01
Pre-experimental (12 *Ss*, 12 stimuli)	.943	<.01	.315	.30
Postexperimental (12 *Ss*, 12 stimuli)	.882	<.01	.644	<.01

* F = Response Frequency
† I% = Percentage of Idiosyncratic Responses.

group. It is apparent that in all cases the index was very high, ranging from .872 to .943. This finding indicates that *Ss* who gave a high frequency of responses to any one stimulus class also tended to give a high frequency of responses to any other stimulus class. This tendency did not seem to be affected by the introduction of the experimental stress variable.

Idiosyncratic Responses

It will be recalled that the percentage of *S*'s responses which that *S*, and no other *S*, gave to a particular stimulus was chosen to represent the popularity factor derived from the factor analytic investigation. In Study I, the relationship between F and I% was found to show a nonsignificant rank-order correlation of +.19. In Study II, however, this relationship, using all 24 *Ss* on the first half of the stimuli, was +.72 ($p - < .01$). This discrepancy led to a further examination of the results of Study I. It was discovered that the low positive correlation could be attributed largely to the behavior of two discrepant *Ss*. These individuals showed the two lowest ranks on frequency of response and the two highest ranks on percentage of idiosyncratic responses. Using only the remaining 18 *Ss*, the rank-order correlation between the two measures increased to +.63 ($p - < .01$). A further investigation of these two *Ss* lent support to the hypothesis that this negative correlation between frequency and percentage of idiosyncratic responses may be diagnostic of deviant adjustment. The following two quotations from the *E*'s postexperimental notes illustrate this point:

> Subject A: "... he said he hoped his answer hadn't been too 'vague' ... his vocabulary is broad but he sometimes couldn't remember what a word meant ... he wants to know more about psychology, about himself and normality ..."
>
> Subject B: "... a bit paranoid, at least suspicious and unsure of self ... might be witty, clever in a hostile way ... probably deep-seated problems and well entrenched defenses ..."

These impressionistic comments tend to bear out the supposition of some pathological processes separating these *Ss* from the rest of the sample. As far as their response behavior was

TABLE 5
PERCENTAGE IDIOSYNCRATIC RESPONSES (I%) FOR THE FOUR STIMULUS CLASSES

Stimulus Class	Total I%		*Ss'* range in I% in Study I
	Study I	Study II	
Nonsense syllables	54%	58%	22%–91%
Adjectives	53%	54%	26%–83%
Drawings	51%	55%	25%–83%
Rorschach cards	62%	57%	37%–83%

concerned, they were very inexpressive individuals, apparently tending to inhibit many of their responses. The resultant behavior was highly idiosyncratic; both *Ss* gave 76 per cent idiosyncratic responses. Comparable individuals either gave a large number of responses with a high percentage of idiosyncratic ones, or else gave few responses, most of which would tend to be given also by other *Ss*.

The range of I% showed large variability. In Study I, for all 20 *Ss*, I% varied from 30 per cent to 76 per cent.[7] The percentage of idiosyncratic responses was surprisingly stable from stimulus class to stimulus class, as shown in Table 5.

Once again there were large individual differences in this variable, and relatively small differences between stimulus classes. As in the frequency variable, it is surprising that the Rorschach stimuli did not appear qualitatively distinct from the other stimulus classes. They elicited somewhat more idiosyncratic material, but only by a few percentage points.

The effect of the experimental conditions on the relative frequency of idiosyncratic responses was similar to the effect on response frequency. An analysis of covariance of I% scores of the control and experimental groups showed a greater increase in the experimental group at the .05 level of significance. When total numbers of idiosyncratic responses were examined, it appeared that the general increase in frequency from the first to the second half differed for the two groups. In the control group, only one per cent of the response increment in the second half of the experiment consisted of idiosyncratic responses, whereas in the experimental group, 91 per cent of the additional responses were idiosyncratic. Thus, it appears that the effect of failure stress was to

[7] See Footnote 2.

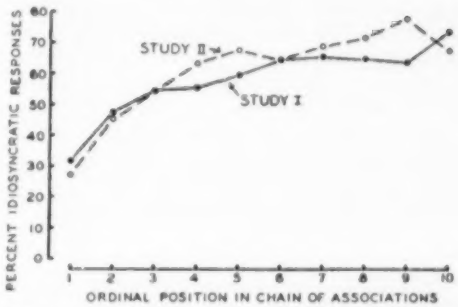

FIG. 1. PERCENTAGE OF IDIOSYNCRATIC RESPONSES ELICITED BY SUCCESSIVE ASSOCIATIONS. EACH POINT REPRESENTS THE TOTAL I% ELICITED BY ALL THE INITIAL, SECOND, THIRD, ETC., RESPONSES GIVEN BY ALL *Ss*

increase the productivity of idiosyncratic material.

Table 4 shows the index of concordance for I% in the various groups. Once again, though not as marked as for the frequency variable, there was a high degree of consistency from stimulus class to stimulus class. The only exception was the pre-experimental group, where consistency was not significant. These same *Ss*, however, seemed to achieve a significant degree of consistency following the failure stress. The nonsignificant finding prior to the stress may be a function of the relatively small number of stimuli used in determining this measure. In any case, the increase in *W* as a function of the experimental variable seems to indicate that *Ss* show greater consistency in their post-stress responses or, in other words, that individual differences are more emphasized as a function of the stress.

The data also provide evidence concerning the degree of idiosyncratic responding as a function of successive associations. Figure 1 shows the total percentage of idiosyncratic responses elicited by each of the first ten associations given to all stimuli. Since there were no differences between subgroups in Study II, the data from all 24 *Ss* were combined. The figure shows that while the initial response to the stimuli was idiosyncratic in about 30 per cent of the cases, this percentage rose steadily to about the 60 per cent level by the fifth or sixth response given. Little increase occurred after that point. These data indicate that while *Ss* on the whole tend to give primarily

nonidiosyncratic material during their first few associations, their later associations tend to be primarily idiosyncratic. However, no increase in the percentage of such material is obtained by the addition of more responses. This suggests that in a word association task, for example, less than 10 responses to any one stimulus are adequate to reach a high level of elicitation of personal, idiosyncratic material.

DISCUSSION

The findings bear out the suggestion that a controlled association situation can be used profitably in describing individual differences in the frequency and idiosyncracy of readily available associative material. The general hypotheses have been substantiated:

1. Differences between *S*s show a wide range of values and constitute a highly significant and consistent source of variance.

2. Individuals show highly consistent modes of responding to a variety of different stimuli. Both the frequency of response and the idiosyncratic content of these responses show repeatable orderings of *S*s across stimulus classes.

3. For the particular stimulus classes used, there are consistent differences in the amount of material these groups of stimuli elicit. While these stimuli were not specifically chosen to provide for a wide range of response variation between classes, they do represent a rather wide variety of visually presentable materials. The particular antecedent learning conditions which make for differentiations between these stimulus groups are of no particular concern to the present investigation. However, the question might be raised whether these differences are pervasive and repeatable under different conditions or whether the particular differences between classes found might be restricted to the particular situation and mode of presentation used in these studies.

4. A failure-stress situation produces measurable differences in response frequency and idiosyncratic content. It appears that the *S*s in Study II responded to the report of failure and the exhortation to do better by producing more associations, and that these additional associations were primarily of an idiosyncratic character. It is, of course, open to question whether this represents a reaction to stress or anxiety per se, or whether it is a function of the particular instructions used. The *S*s' immediate reaction may well have been a conscious attempt specifically to produce more idiosyncratic responses. Taking part in a psychological experiment may well imply that one is expected to produce personal associations, while failure in a free responding situation is easily interpretable as a criticism of slow or inadequate response rate.

It has been indicated above that the Rorschach cards did not, on the basis of the measures used, represent a qualitatively different set of stimuli. Only on the percentage of idiosyncratic responses did they show a slight tendency to elicit more such materials than the other stimulus classes. This limited finding need not constitute an indictment of inkblots as useful stimulus material. The response categories utilized are extremely limited, and an extended content analysis of the material elicited by the different stimulus classes might indicate some advantage from the responses given to the Rorschach cards in terms of personality theory concepts. An impressionistic survey of the responses, however, does not substantiate this hypothesis at first glance. Rather, it appears that if the responses to the Rorschach cards tend to be useful in that sense, the *S* is also likely to give similar content to the other stimuli. This does suggest that diagnostic devices might well include more mundane stimulus materials than the inkblots.

The finding of deviant behavior in the cases of two *S*s in Study I suggests that the present method might be refined for diagnostic purposes. Deviance in terms of idiosyncratic thought processes, such as in schizophrenia, might be predicted to show just such high idiosyncratic response frequencies coupled with low rates of responding as was found in these cases. However, such speculation must await further studies of controlled associations in particular pathological populations.

In these studies the response variables have not been related to any other personality measures. The main reason is that this investigation was considered primarily exploratory in character. Some theoretical notions both from experimental and personality studies have suggested that the associations freely available to a subject are theoretically and pragmatically useful in describing other aspects of his behavior. These studies have shown in a limited

context that these responses prove to be consistent and measurable aspects of individual differences. Rather than relate them to other personality measures, when a rationale for such relationships is lacking, it is suggested that further investigation of the antecedent conditions and behavioral correlates of these response measures would be more relevant and in the long run more useful.

SUMMARY

Two studies were conducted to test five general hypotheses about the frequency and content of verbal associations elicited during a 30-sec. period. In the first study, 20 *Ss* responded to 40 stimuli, 10 each from four stimulus classes: Nonsense Syllables, Adjectives, Line Drawings, and Rorschach cards. In the second study, the effect of a failure-stress situation on response frequency and content was tested with a total of 24 *Ss*.

Two variables, the frequency of response and the idiosyncracy of response (defined as the percentage of individual-specific responses) were selected for detailed study on the basis of a factor analysis of 13 response variables. The main findings were as follows:

1. Differences between *Ss* in frequency and idiosyncracy showed a wide range of values. Individual differences in response frequency contributed the most significant source of variation.

2. There were highly significant consistencies in frequency and idiosyncracy within *Ss* across stimulus classes.

3. There were significant differences between stimulus classes in the frequency of responses elicited. Variability between stimulus classes was significantly greater than within stimulus classes.

4. The failure-stress situation tended to increase the productivity of idiosyncratic material.

The results indicate the utility of "controlled" associations as an individual difference variable and their possible use as a diagnostic tool.

REFERENCES

1. HEIDBREDER, EDNA. The attainment of concepts: I. Terminology and methodology. *J. gen. Psychol.*, 1946, **35**, 173–189.
2. KIMBLE, G. A., & DUFORT, R. H. Meaningfulness and isolation as factors in verbal learning. *J. exp. Psychol.*, 1955, **50**, 361–368.
3. LAZARUS, R. S., DEESE, J., & OSLER, S. F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.
4. MANDLER, G. Response factors in human learning. *Psychol. Rev.*, 1954, **61**, 235–244.
5. MANDLER, G. Perceptual learning: Stimulus differentiation of differentiating response. *Psychol. Rep.*, 1955, **1**, 79–82.
6. MANDLER, G. Associative frequency and associative prepotency as measures of response to nonsense syllables. *Amer. J. Psychol.*, 1955, **68**, 662–665.
7. MANDLER, G., & HUTTENLOCHER, JANELLEN. The relationship between associative frequency, associative ability, and paired-associate learning. *Amer. J. Psychol.*, 1956, **69**, 424–428.
8. NOBLE, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.
9. NOBLE, C. E. The role of stimulus meaning (*m*) in serial verbal learning. *J. exp. Psychol.*, 1952, **43**, 437–446; 1952, **44**, 465.
10. NOBLE, C. E. The meaning-familiarity relationship. *Psychol. Rev.*, 1953, **60**, 89–98.

Received August 27, 1956.

THE RELATION OF EGO CONTROL TO OVERT AGGRESSION AND DEPENDENCY

NORMAN LIVSON AND PAUL H. MUSSEN[1]

University of California

RECENT research indicates that aggression and dependency are, in large part, learned drives and that individual differences in strength of these drives are related to early learning experiences. Thus, the strength of the young child's dependency drive has been shown to be a consequent of amount of feeding frustration during infancy and degree of maternal nurturance and punishment during the preschool years (10), while maternal discipline and the extent of the child's identification with an aggressive like-sexed parent appear to be related to the strength of the child's aggressive drive (8, 10).

The amount of overt aggression and dependence the child manifests, however, is a function not only of the strength of his drives but also of the strength of the inhibitions against the expression of these drives. In the course of becoming socialized, i.e., acquiring the culturally customary and accepted forms of behavior, the child learns to conform with the taboos against the expression of anger, sexual impulses, and tendencies to be dependent (5). Individual differences among children in strength of inhibition of these impulses have been shown to be related to social class membership. For example, in the American middle class, strong pressures for the inhibition of aggressive and sexual behavior are applied relatively early in the child's life (6, 11). Independence training in this class is also initiated when the child is very young, although it is completed at a late age. Thus, compared with lower class children, those of the middle class are more likely to withhold overt manifestations of aggression, sexuality, and dependency. Knowledge of class differences, however, tells us little about how the child develops control of his drives. To understand this process, it is necessary to focus upon the individual experiences of the child.

According to psychoanalytic theory, the internalization of relatively strong inhibitions against the expression of these impulses may be attributed to the functioning of the ego, the organized, executive subdivision of personality. This theory maintains that, as the child becomes aware of the reality principle and as his ego begins to develop, he acquires a rudimentary ability to test reality, to tolerate or bind tensions and to postpone gratifications. These abilities, which are major aspects or components of ego-control capacity or ego strength, presumably enable the individual to control his impulses and to cope adequately with frustrations.

With further development of ego-control capacity, there is increased recognition of, and adjustment to, culturally defined reality. Social drives become relatively more influential, compared with primary drives and immediate impulses, in guiding the child's perception of the social environment and in directing his social behavior. A child who acquires a high degree of ego-control capacity is likely to manifest more socialized behavior—behavior which conforms with socially accepted standards—than a child whose ego-control capacity is less highly developed.

Based on these theoretical considerations, the present investigation was designed to test the general hypothesis that individual differences in ego-control capacity are related to degrees of inhibition of aggression and dependency behavior.

It was first necessary to obtain measures of generalized ego-control capacity sufficiently sensitive and objective to provide reliable quantitative indices of individual differences in ego control in a group of normal children. Block and Martin (3) have presented two experimental techniques that elicit behaviors which permit assessment of two important facets of ego-control capacity. They obtained measures of nursery school children's abilities to delay gratification and to bind tension in two procedures. These measures were related to children's tendencies to regress and to attack the barrier in a replication of the Barker, Dembo, and Lewin experiment (2). In general, the results supported their hypotheses

[1] The authors wish to acknowledge their gratitude to Dr. Ernest Haggard for his helpful criticism of this manuscript.

that lack of ego control in children is associated both with a decrement in play constructiveness and with a tendency to attack the frustrating barrier directly.

The rationale underlying Block and Martin's use of the delay of gratification and cosatiation techniques stemmed from Lewinian conceptions, specifically those dealing with permeability of boundaries between need systems. However, it is also possible to view both tasks as frustration procedures and, hence, to consider the individual's performance as an index of his frustration tolerance—a basic aspect of his ego-control capacity. The reasoning underlying this formulation seems immediately apparent for delay of gratification; for behavior in the cosatiation task, it draws upon Burton's (4) finding that satiation may be psychologically the equivalent of frustration. Whatever the rationale for the use of these techniques, if they yield indices that provide valid measures of general internalized ego control, they should relate to the child's ability to inhibit the *overt* expression of socially disapproved drives in everyday life.

Hypotheses

We propose to test the following two specific hypotheses in this study:

1. Among middle class children, there is a negative relationship between children's ego-control capacities and the amount of overt aggression they express in the nursery school situation.

2. In this group, there is also a negative relationship between ego-control capacity and degree of overt dependent behavior in nursery school children.

METHOD

Subjects

Subjects (*Ss*) were 36 children, 19 boys and 17 girls, ranging in age from 38 to 59 months. Of this group, the three-year-olds attended the morning session, and the four-year-olds the afternoon session, of the nursery school at the Institute of Child Welfare of the University of California. The sample is a very select one: almost all of the children come from highly educated, upper-middle class families.

Measures of Ego-Control[2]

The two techniques for measuring ego control, cosatiation and delay of gratification, were adopted with slight modifications from Block and Martin (3).

[2] We are indebted to Dr. Frances Orr for lending the apparatus for these procedures.

The first step in the cosatiation procedure was to show the *S* how a number of spools could be packed in a rectangular rack on top of a box and then be made to fall into the box by sliding the rack. After the spools had fallen, the rack could be refilled and the same procedure repeated. The child was then told that he could play the game as long as he wished and he began operating the sliding rack himself. As soon as he indicated that he did not want to perform the task anymore, i.e., when he was satiated, the experimenter (*E*) recorded the number of trials of spool packing. *E* then placed colored blocks into the rack, telling the child he could play with these, as he had with spools, as long as he wished. The number of trials for this second performance was recorded.

The extent of deterioration in performance on the second task, following satiation on the prior one, may be expressed by an index of cosatiation very similar to the one used by Kounin (7). Performance on the first task serves as the basis of comparison: trials to satiation on the first task (spools) minus trials to satiation on the second (blocks) is divided by trials to satiation on the first task. This was the Cosatiation score. A high Cosatiation index is regarded as indicating relatively low ego control, i.e., satiation in the first task resulted in poorer performance in the second.

The experiment on delay of gratification permits direct quantitative assessment of the child's ability to delay immediate gratification of needs in order to attain greater rewards in the future. At the beginning of the experiment, *E* showed the child a small bowl containing a variety of small toys (commercially marked as bracelet "charms"). The *S* was then shown a bag full of small packages wrapped in wax paper and was told that each package contained a toy like those displayed in the bowl. It was explained that he would be given all the toys he accumulated. However, once stopped, he could not begin again.

The apparatus and procedure for acquiring the toys was then demonstrated to the child. The apparatus was a toy known as a "coke loader" which consists of a crank-operated cart that could be moved up an incline until it tipped to empty the contents. The procedure was to put one toy at a time into the cart and allow *S* to operate the crank until the toy fell into a box. After insuring that *S* understood how the toy worked, *E* again told him that he could win as many toys as he liked and, at his request, could stop playing and be given all the toys in the box. If the child had not voluntarily stopped by the time he had accumulated 20 toys, *E* ended the session by saying that time was up.

It was assumed that the child was in a conflict situation during this procedure; that he would want to claim his packages quickly to discover what they contained and to play with the toys, while at the same time, that he would want to get more toys. Presumably, the child who generally has a greater capacity to delay gratification would be able to control his desires to have his packages immediately and to examine their contents. Hence he would be expected to accumulate a relatively large number of toys before deciding to stop. The delay of gratification (Delay) score was simply the number of packages accumulated in the box.

Measures of Overt Aggression and Dependency[3]

These were based on reports of observations of children's nursery school behavior by the three teachers in each (morning and afternoon) session. Twelve children in each session were selected for special observation for a period of two weeks or ten school days. This procedure was repeated for two more groups of 12 in a subsequent two-week session. Of the total of 48 children observed, 36 were tested in both ego-control situations, and comprise the sample of this study.

Two types of records were made: daily behavior reports for aggression and dependency, and weekly rankings of the 12 children being observed with respect to these two characteristics. The daily behavior report for aggression consisted of a checklist of 17 items of aggressive behavior. Included were: interfering with others' activities; teasing or insulting; threatening; quarrelling, screaming or swearing; physically attacking; throwing objects; destroying something intentionally; disobeying; temper tantrums. For dependency behavior there was a twelve-item checklist that included seeking physical contact, asking for help, seeking attention and/or approval, needing security object, wanting to go home, revealing hurt feelings, "helpless" crying, and asking permission unnecessarily. Each teacher filled out these two checklists for each child observed on that day.

The incidence of aggressive and dependent behavior as indicated by the total number of items checked by one or more teachers on each day was tabulated for each child over the entire observation period. Unfortunately, as a result of children's absences from school, not all *Ss* were observed for the same number of days; consequently, the number of daily behavior reports available was not the same for all *Ss*. In order to obtain comparable measures, the total incidence of aggressive and dependent behaviors for each *S* was divided by the number of days he was observed, yielding aggressive Acts/Day and dependent Acts/Day scores. These scores were not computed for the three *Ss* who were absent from school for more than five of the ten observation days. The remaining 33 *Ss* had aggression scores ranging from 0 (least) to 4.3 (most), with a median of .30; while the dependent Acts/Day scores ranged from 0 to 3.4, with a median of .65.

At the end of each week of observation, each of the three teachers independently ranked the 12 children she had been observing that week with respect to overall overt expression of aggression and dependency. Thus, there were six weekly rankings on each trait for each child which were summed to obtain his Pooled Ranks scores for each of the two behaviors.

The interrater reliability for the Pooled Ranks scores was determined in the following way. For each trait, each teacher-observer's average rank for the child, based on the two weekly rankings she assigned him, was treated as a score. The resulting scores of the morning and afternoon groups were then ranked separately. The rank-difference correlations between each teacher in a session and the other two teachers in that session were computed. The median coefficient for aggression (based on six coefficients, three from each of the two sets of observers) was .80, for dependency, .48. In order to increase the low reliability of the dependence Pooled Ranks score, the analysis excluded those cases about which the observers showed the most disagreement. Specifically, if the average ranks on dependency assigned to *S* by any pair of observers differed by more than three units, no score was computed for him. This procedure reduced the number of Pooled Rank scores for dependence from 33 to 22.

[3] We wish to thank E. Peters and H. Tyau, head teachers, and E. Benjamin, E. Boydston, E. Elson, and H. Shimomaki, nursery school assistants, for their cooperation in this aspect of the study.

RESULTS AND DISCUSSION

In order to test the two major hypotheses, the measures of ego control were related to the indices of overt aggression and dependence. Since the distributions of scores on Delay and Cosatiation were non-normal and included a relatively large number of tied scores, the sample was dichotomized into High and Low ego-control groups on each of the measures, the dichotomization point in each case being selected so as to provide as nearly equal numbers as possible in the two groups. (Since High ego control is indicated by a *low* score in Cosatiation and a *high* score in Delay, the terms High and Low will henceforth be used to refer only to ego-control status.) For Cosatiation (range, −.20 to .10; median, zero) the High ego-control group (N = 19) thus constituted included scores of zero and below; the Low group (N = 17), scores above zero. On the Delay measure, the 20 *Ss* who achieved the maximum score of 20 trials comprised the High group, while the Low group (N = 16) included the remaining *Ss* whose scores ranged from 5 to 18.

Composite ego-control categories were devised in order to test the combined predictive ability of the two ego-control measures which tended to be positively, though not significantly, related (P = .20 by a U test comparing the actual Cosatiation scores of the High and Low Delay groups). The High Composite group (N = 11) included those *Ss* who were classified as High ego control on both Cosatiation and Delay; the Low Composite group (N = 8) was composed of *Ss* assigned to the Low groups on both measures.

In view of the dichotomization of the ego-control measures and the non-normality of all four of the behavior score total distributions, the U test (9) was employed to compare rank

transformations of all overt behavior scores of the Highs and Lows on each measure of ego control. (Figure 1 summarizes all these data and shows each *S*'s ego-control status—High or Low—as derived from the two experimental measures and his aggression and dependency scores. The legend in upper right portion of Fig. 1 indicates ego-control status on both Cosatiation and Delay measures, *not* actual scores. "H" and "L" in key refer to ego-control level on the two measures. Orientation of the Acts/Day and Pooled Ranks scales are the same; in both cases, the upper end of the scale denotes greater amount of the behavior. Reference to the appropriate portions of this figure will serve to illustrate the nature of the comparison made by a given *U* test.) The results

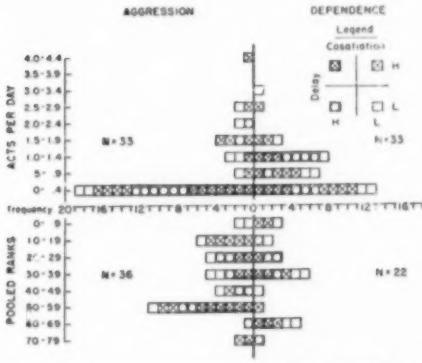

FIG. 1. DISTRIBUTIONS OF AGGRESSION AND DEPENDENCY SCORES FOR HIGH AND LOW EGO-CONTROL GROUPS. EACH UNIT IN HISTOGRAM REPRESENTS an *S*. (See text for explanation of legend.)

of these tests and their significance levels for the experimental hypotheses are summarized in Table 1.

Compared with *S*s high in ego control, as measured by the Cosatiation score, those low in this measure of ego control scored significantly higher on both indices of overt aggression. There was also some tendency for those who showed relatively high ability to delay gratification to express fewer aggressive Acts/Day. For Pooled Ranks-Aggression, the difference between High and Low ego-control groups on Delay was in the same direction, though the difference was not significant. When the two ego-control measures were combined into the Composite score, the less controlled *S*s were found to be significantly more aggressive than the more highly controlled *S*s. Thus, there is general support of the first hypothesis, that ego-control capacity is positively related to ability to withhold expression of aggression.

Since the data from children of both sexes and of two age groups were combined in the analyses, these relationships might be attributable to some extent to covariation of the measures of ego control and aggression with sex and/or age. Thus, it is possible that girls develop greater ego-control capacity earlier and are also less likely to express aggression overtly. Similarly, both ego control and inhibition of overt aggression may increase with age. In our sample, girls did show significantly fewer aggressive Acts/Day but were not significantly different from boys on either of the ego-control measures. None of the measures of aggression or ego control showed significant relationships

TABLE 1

COMPARISONS OF DISTRIBUTIONS OF AGGRESSION AND DEPENDENCE SCORES BETWEEN HIGH AND LOW EGO-CONTROL GROUPS

Behavior	Measure	Ego-Control Measure								
		Cosatiation			Delay			Composite		
		*N**	*U***	*p****	*N*	*U*	*P*	*N*	*U*	*P*
Aggression	Acts/Day	33 (17, 16)	86.5	.04	33 (18, 15)	105.5	.14	16 (9, 7)	12.5	.02
	Pooled Ranks	36 (19, 17)	82.5	< .01	36 (20, 16)	138.5	> .20	19 (11, 8)	17.0	.01
Dependence	Acts/Day	33 (17, 16)	127.0	> .20	33 (18, 15)	111.5	.20	16 (9, 7)	24.0	> .20
	Pooled Ranks	22 (10, 12)	54.5	> .20	22 (10, 12)	58.5	> .20	12 (5, 7)	13.0	> .20

* First number is the total *N* in the comparison; numbers in parentheses refer to the partitioning of this total into the *N*s for the High and Low ego-control groups, respectively.

** All *U* values in this table are less than the value to be expected by the null hypothesis (one half the product of the two subgroups *N*s) and are in the predicted direction.

*** All *P* values are based on one-tailed tests as is required by the hypotheses.

with age. Hence, partialling out sex or age should not attenuate the reported relationships.

The data did not support the second hypothesis concerning the relationship between ego control and the overt expression of dependency needs. Although all six comparisons between High and Low groups on the three measures of ego control yielded differences in the predicted direction, none were statistically significant. This failure to detect relationships between ego-control and dependency may be due to several factors. It is quite possible, in the first place, that dependency is inherently difficult to observe and measure under conditions of free observation in a nursery school situation. The original low correlations among the observers on the dependency rankings attest to this difficulty. Another factor to be considered is that these children come from a highly select and homogeneous group of upper-middle class, highly educated families, a very large proportion of the parents having advanced academic or professional degrees. In such a group "getting along with others" is an important social value and is often considered a criterion of adequate social adjustment. Hence, socialization of aggressive impulses, especially in interpersonal situations, is likely to be emphasized from early childhood. While children in this group may be encouraged to acquire independence gradually, these parents are also likely to be permissive with respect to the expression of dependency for a relatively long period of time. Consequently, the ego-control capacity of the preschool child may have little bearing upon the amount of dependent behavior he manifests.

Our data offer some indirect evidence to support the hypothesis that there is less emphasis on the control of dependence than of aggression in this sample. Examination of the frequency distributions for the Acts/Day scores in Figure 1 shows that the incidence of both aggression and dependence generate J curves which Allport (1) has demonstrated to be typical for behavior in a number of conformity-demanding situations. He has reasoned that the steeper the slope at the high conformity end of a scale, the stronger are the conformity-producing pressures influencing the particular behavior. Dependence exhibits a more gradual slope than does aggression, with the relative number of very high conformists being greater for aggressive Acts/Day. This difference may be quantified by a comparison of the skewness of the two distributions; the skewness indices (3 [Mean-Median]/Standard Deviation) are 2.78 for aggression and only 1.06 for dependence. This difference in skewness is consistent with the hypothesis of differential conformity pressures for the control of these two drives.

The data indicating that ego-control capacity is positively related to the inhibition of aggression in interpersonal relationships are consistent with the finding of Block and Martin (3) that children high in ego control make relatively few direct attacks on a frustrating barrier in an experimental situation. Thus, it is possible to interpret the findings of the present study—as Block and Martin viewed their results—as generally supportive of a Lewinian theoretical formulation.

It is also of interest to note that in both studies, performance in the cosatiation situation proved to be a better predicter than did delay of gratification scores. We can only suggest that the Cosatiation score may have been a purer, less confounded index of ego control than Delay. It measured the child's ability to continue to perform without extrinsic rewards even though he had just been satiated and was free to stop whenever he wished. Such behavior seems to correspond closely to our definition of internalized ego control. In contrast, since the delay situation provided extrinsic rewards, the final Delay score may have represented a confounding of internalized ego-control factors and a desire to obtain the available rewards. Thus, a High ego-control score on this task may have been due either to ability to delay gratification, to high motivation to possess many toys, or to both.

The results of the present study extend the meaning and implications of the concept of ego control by demonstrating how it may be useful in understanding and predicting one aspect of socialization, the inhibition of aggressive expression. Furthermore, some of the findings suggest that the concept has predictive utility only for behavior to which socialization pressure has been applied. Thus the absence of a significant relationship between ego-control measures and overt dependency in these data

may be due to the fact that there is less emphasis on dependency training in this group of *S*s than we had initially assumed. For similar reasons, we would not expect to find relationship between ego control and the inhibition of overt aggression in preschool children from a social class group in which the expression of overt aggression is not discouraged at this age. Conversely, in groups which stress early training in the suppression of overt dependency behavior, our original hypothesis of negative correlation between ego-control capacity and the incidence of dependency manifestations should be confirmed.

SUMMARY

This study tests two hypotheses concerning the relationships between ego-control capacity in children and their ability to inhibit the overt expression of aggression and dependency needs. Thirty-six nursery school children were scored on indices of aggressive and dependent behavior manifested during a two-week observation period. These scores were related to two measures of ego control derived from performance in a delay of gratification and in a cosatiation situation.

There is a significant negative relationship between amount of expressed aggression and ego-control as assessed by the cosatiation measure. The delay measure is also negatively related to aggression, though not significantly. A composite ego-control score, involving both measures, was also related to aggression.

Dependent behavior is not predictable from any of the ego-control indices. The argument is made that ego-control capacity should be expected to relate to the degree of expression of a drive only when that drive has been subjected to socialization pressures.

REFERENCES

1. ALLPORT, F. H. The J-curve hypothesis of conforming behavior. *J. soc. Psychol.*, 1934, **5**, 141–183.
2. BARKER, R. G., DEMBO, TAMARA, & LEWIN, K. Frustration and regression: an experiment with young children. *Univer. Iowa Stud. Child Welf.*, 1941, **18**, Whole No. 1.
3. BLOCK, JEANNE, & MARTIN, B. Prediction of behavior of children under frustration. *J. abnorm. soc. Psychol.*, 1955, **51**, 281–285.
4. BURTON, A. The aggression of young children following satiation. *Amer. J. Orthopsychiat.*, 1942, **12**, 262–267.
5. CHILD, I. L. Socialization. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge: Addison-Wesley, 1954. Pp. 665–692.
6. DAVIS, A., & HAVIGHURST, R. J. Social class and color differences in child-rearing. *Amer. sociol. Rev.*, 1946, **11**, 698–710.
7. KOUNIN, J. S. Experimental studies of rigidity I. *Charact. Pers.*, 1941, **9**, 251–272.
8. LEVIN, H., & SEARS, R. R. Identification with parents as a determinant of doll play aggression. *Child Develpm.*, 1956, **27**, 135–153.
9. MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.
10. SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, P. S. Some child-rearing antecedents of aggression and dependency in young children. *Genet. psychol. Monogr.*, 1953, **47**, 135–234.
11. WHITING, J. W. M., & CHILD, I. L. *Child training and personality*. New Haven: Yale Univer. Press, 1953.

Received August 30, 1956.

ATTITUDE CHANGE THROUGH REWARD OF VERBAL BEHAVIOR[1]

WILLIAM ABBOTT SCOTT

University of Colorado

PUBLIC opinion researchers are familiar, through anecdote if not personal acquaintance, with the woman who, when asked her opinion on social security, replied that she didn't know because no one had ever asked her, and she never knew what her opinions were until she opened her mouth and expressed them. Perhaps this apocryphal respondent was displaying more insight into the psychological processes underlying her attitude than is readily apparent. Though cognitive approaches to attitude development are more flattering in their emphasis on the rationally cognized relation of an object to the person's goals (1, 3, 8, 9), nonrational determiners may be equally important in the acquisition of attitudes.

According to Doob (4) an attitude may be regarded, like a habit, as an implicit anticipatory response which mediates overt behaviors, and arises out of them through response reinforcement. Thus, it is conceivable that an opinion be expressed initially, in the absence of a supporting attitude, but if the verbal behavior is rewarded, the corresponding attitude may develop and mediate subsequent opinion expressions in the presence of similar cues. This response reinforcement formulation thus focuses more on reward of verbal expressions than on cognized means-end relationships as critical factors in attitude development.

The two formulations, and the sets of variables they employ, are by no means incompatible. It is reasonable to suggest, for instance, that lacking a well structured and salient set of cognitive associations regarding an object, a person may be more readily induced to respond to it in nonlogical fashion, the outcome of his response then contributing to a nascent cognitive structure; but given a set of values and cognitions that can readily be applied to a new object, these will influence the person's attitudinal set and thus produce a response consistent with the pre-existing cognitive structure. It is not the purpose of this paper to propose a systematically inclusive formulation of these two classes of determinants, but rather to offer evidence concerning the efficacy of response reinforcement in the determination of attitudes.

Other studies (2, 5, 10) have demonstrated the influence of social approval and disapproval on the modification of verbal behavior, but they provide no information about their effects on the underlying attitudes. The present study was designed to explore, in a controlled fashion, the effects on attitudes of reward and punishment of expressed opinions. Specifically, it was predicted that subjects (*S*s) rewarded by group approval for expressing opinions opposite from their initial attitudes would show a change in the direction of the expressed opinions, while *S*s punished by group disapproval for expressing contrary opinions would not show such a change. This prediction implies that reward of a new behavior increases the relative strength of the underlying predisposition, while punishment of a new behavior leads the *S* to revert to his formerly preferred response disposition.

METHOD

All students in 29 general psychology discussion sections were administered free-response questions on mimeographed sheets to assess their attitudes toward three controversial issues:

1. *Universal military training.* It has been proposed that all physically able males between the ages of 18 and 25 be required to spend two years in the armed forces so that they will be trained and ready for armed service in the event of a national emergency. How do you feel about this proposal?

2. *Night hours for women students.* Senior women at the University are not required to be back in their dorms or houses by any set time at night. Some students are requesting that all women be given the same freedom of hours. How do you feel about this?

3. *De-emphasis of football.* It has been suggested that too much emphasis is placed on football at the University, in scholarships and special treatment for players. How do you feel about this?

Each of the classes was assigned, in systematic fashion, one of the three issues for debate. The instructor read the class members' opinions on the designated

[1] The author is indebted to Prof. Maurice P. Smith for his suggestions regarding design and report of the study, and to the following graduate teaching fellows who assisted in the collection of data: H. Hess, P. Khanna, A. Mathews, and I. Richardson.

issue and selected from each class two which appeared definitely favorable and two definitely unfavorable. The four students were contacted outside of class and their cooperation requested in an "experiment to see how much they could affect the opinions of class members" by debating the particular issue; but each student would take the side opposite to his own opinion. Two pairs of debators were formed for each class, members with "pro" opinions taking the "con" side of the argument, and vice versa. It was explained that, following the debate, the quality of performance would be assessed in two ways: first, by a class vote on which member of each pair did the better job; and second, by a retest of the class' opinions on the issue, "to see in which direction opinions are influenced."

The debators were counseled to avoid mentioning to anyone that they didn't really believe what they were to say. (This was intended to help prevent the occurrence of interfering responses (see 7) which might negate the effect of the experimental stimulus.) In pairing the debators, an attempt was made to equate their verbal abilities, as evidenced in their written opinions, in order that a subsequent false report of the vote would appear valid.

The debates were held two weeks after the original assessment of attitudes. One pair of opponents was called into the classroom, while the other pair waited in an adjoining room. The "pro" debator spoke first, for three minutes, followed by a three-minute presentation from the "con" debator. Then each had a two-minute rebuttal period, in reverse order. The first pair left the room, and the second pair of opponents entered for a similar performance. Then, with all four debators present, voting instructions were given to the class ("On the first ballot write which side—pro or con—did a better job on the first debate. On the second ballot write which side—pro or con—did a better job on the second debate").

The ballots were handed in, and the instructor made a pretense of counting them, while one of the students tallied votes (as read by the instructor) on the blackboard, under the name of each debator. It had been determined in advance for each pair of debators in all classes who would "win." "Winners" and "losers" were assigned alternately through the pairs of debators, so that half the time the "pros" won, and half the time the "cons" won.

The instructor then repeated the implications of the vote, mentioning the names of the "winners" and "losers," and explained that the next step in the judgment would be to see how the class's attitudes had been affected. All class members were given sheets of paper on which were printed the statement of the issue that had been debated—exactly like the first opinion assessment, except that only one issue appeared on the sheet. Debators were also asked to write their present opinions, "to see if there was any change in them."

After the opinion retest forms were collected, the nature and purpose of the experiment were explained, the pre- and postopinions of the debators were read, and an attempt made to reduce the threat of the situation for the "losers" and for those debators whose opinions had ostensibly shifted.

The pre- and postopinions of the debators were subsequently typed on cards, after the deletion of certain passages that might indicate which were postopinions (such as "I still feel that . . ."). The cards were identified only by numbers, duplicated on the original opinion questionnaires, and the numbers assigned in "random" fashion, so as to prevent contamination in coding. Opinions on the cards were coded on a seven-point scale: 1—very pro; 2—pro; 3—pro, qualified; 4—neutral or balanced pro and con; 5—con, qualified; 6—con; 7—very con. Then, by reference to the original questionnaires, each card was identified as a pre- or postopinion of a "winner" or "loser."

Of the 58 pairs of debators initially designated, only 36 finally produced complete data. Twenty-one pairs failed to debate, due either to absences or to rescheduling of classes; the postopinion of one member of the final pair was not codeable. The 36 debates took place in 19 different classes, with six different instructors. From these classes a sample of 36 students whose original opinions were definitely pro or con, but who did not debate, was selected as a control group. Their pre- and postopinions were similarly transferred to cards with random indentification by numbers, and coded on the same seven-point scale.

RESULTS

It was predicted that debators who "won" would show an average change in the direction of their debate greater than the corresponding change shown by "losers" and greater than the average change of the control group. Data bearing on this prediction are presented in Table 1. Winners changed an average of 1.25 points (in the direction of debate) on the seven-point scale, while losers changed an average of 0.17 points in the opposite direction (more extreme than their original positions). The control group changed an average of 0.31

TABLE 1

CHANGES OF WINNERS, LOSERS, AND CONTROLS FROM PRE- TO POSTTEST

(Positive sign indicates mean change in direction opposite to *Ss'* original opinions)

	Winners (N = 36)	Losers (N = 36)	Controls (N = 36)
Number of changers (either direction)	24	16	18
Number of changers (toward opp. direction)	21	7	11
Mean change	+1.25	−0.17	+0.31
SD of change	1.77	1.17	1.37

Difference in mean changes:
Winners *vs.* losers: $t = 3.97$; $p < .001$.
Winners *vs.* controls: $t = 2.49$; $p < .02$.
Losers *vs.* controls. $t = -1.58$; $p > .10$.

points toward the opinion opposite from their original one.

The difference between mean changes of winners and losers is significant beyond the .001 level of confidence, and the difference between winners and controls is significant beyond the .02 level of confidence. The difference between losers and controls is not significant.

An analysis of mean changes for the three issues separately, and for original pros versus original cons showed no significant differences among these groups, so the prediction was confirmed regardless of the issue and regardless of which extreme opinion the debator started with. When the *proportions of changers* in the three groups are compared, results are consistent with those for mean changes. More winning individuals shifted toward the opposite opinion than either losers or controls.

DISCUSSION

The experiment was designed to give both winners and losers equivalent contact with arguments for the opposite side, and to give both the experience of verbalizing these arguments as if they were their own. What distinguished the two groups was the brief experience of either group approval or group disapproval for their performances. These performances involved principally the oral presentation of arguments—verbal behavior to which they were unaccustomed. The vote of "win" is presumed to have reinforced the verbal behavior and with it the accompanying implicit responses—attitudes and cognitive support for them. The vote of "lose" presumably weakened whatever response tendencies had been established by the overt behavior or by cognitive contact with the opposite side, so that *S*s reverted to their pre-existing attitudes.

If the opinion posttests be regarded as valid reflections of *S*s' attitudes following the experimental stimulus, then one may conclude that social reward for expressing a new opinion tends to reinforce the concomitant attitudes, whereas social disapproval of the new behavior tends to lead to nonreinforcement of the accompanying attitudes. It appears that, in the present experiment, the brief experience of reward or punishment was a significant factor in accounting for attitude change, rather than the more prolonged experience of cognitive contact with opposing arguments.

This finding is, however, subject to at least two interpretations. It may be that reward of a new behavior was the crucial factor, while cognitive contact in and of itself had no effect. Or it may be that contact with opposing arguments produced a change in attitude, which was then reinforced or extinguished depending on whether the first overt expressions were rewarded or punished. Since Janis and King (6) showed that contact with the opposite side tended, by itself, to produce some change, the latter interpretation appears more plausible. In order to help determine the relative importance of these two influences—cognitive contact and reinforcement—in this situation, it would have been necessary to have some *S*s debate without a subsequent "winning" or "losing" experience. This experiment did not provide for such a condition. Regardless of whether one accepts a single-factor or a two-factor interpretation of the results, it appears that they speak clearly for the significance of reward in producing attitude change in this type of situation.

It should be noted that very few of the changes in opinion were of a spectacular sort. (Only seven of the rewarded *S*s reversed the direction of their opinions; the other fourteen changers merely showed some weakening of their initial opinions, as indicated by a shift along the seven-point scale toward the neutral position.) Although the experience of success or failure was made as strong as possible under these restricted conditions, it was nevertheless a brief and perhaps not too potent one. Between the announcement of the "class vote" and the posttest of opinion no more than five minutes elapsed.

The degree to which the new opinions persisted was not assessed in this study. It is reasonable to suppose that effects of the experimental situation were transitory, since the *S*s' initial attitudes presumably found considerable social support in their current friendship groups. In order to produce enduring attitude changes, it would probably be necessary to so change the *S*s' social environments that the new behaviors would receive continued reinforcement. But transitory or enduring, the principles of attitude change illustrated in this experiment would appear to be applicable.

SUMMARY

An experiment in attitude change was suggested by Doob's learning theory of attitude development. Seventy-two *Ss* were induced to engage in debates on three different issues, taking sides opposite to those which they had indicated as their own in an opinion pretest. Half of the *Ss* were rewarded, in predetermined order, by a purported vote which proclaimed them the better debators, while the other half were punished by presumably losing the debate. Posttests of *Ss*' opinions showed a tendency of the "winners" to change their opinions in the direction of their debates, while the "losers" did not change significantly. A control group of nondebators likewise showed no significant change in opinions.

REFERENCES

1. CARLSON, E. Attitude change through modification of attitude structure. *J. abnorm. soc. Psychol.*, 1956, **52**, 256–261.
2. COHEN, B. D. *et al.* Experimental manipulation of verbal behavior. *J. exp. Psychol.*, 1954, **47**, 106–110.
3. CROCKETT, W. H. Attitude change as a function of cognitive differentiation and affect under conditions of norm-presentation with and without counter-arguments. Unpublished doctor's dissertation, Univer. of Michigan, 1952.
4. DOOB, L. W. The behavior of attitudes. *Psychol. Rev.*, 1947, **54**, 135–156.
5. HILDRUM, D. C., & BROWN, R. W. Verbal reinforcement and interviewer bias. *J. abnorm. soc. Psychol.*, 1956, **53**, 108–111.
6. JANIS, I. L., & KING, B. T., The influence of role playing on opinion change. *J. abnorm. soc. Psychol.*, 1954, **49**, 211–218.
7. KELMAN, H. C. Attitude change as a function of response restriction. *Hum. Rel.*, 1953, **6**, 185–214.
8. KRECH, D., & CRUTCHFIELD, R. S. *Theory and problems of social psychology.* New York.: McGraw-Hill, 1948. Chaps. 5–6.
9. ROSENBERG, M. J. Cognitive structure and attitudinal affect. *J. abnorm. soc. Psychol.*, 1956, **53**, 367–372.
10. VERPLANCK, W. S. The control of the content of conversation: reinforcement of statements of opinion. *J. abnorm. soc. Psychol.*, 1955, **51**, 668–676.

Received August 30, 1956.

IDENTIFICATION AND OBSERVATIONAL LEARNING FROM FILMS[1]

ELEANOR E. MACCOBY AND WILLIAM CODY WILSON
Harvard University

IN the studies reported here, our interest is focused upon the process of "passive" learning. We wish to explore the conditions which govern the nature and extent of a viewer's learning the contents of a dramatic production which he views in a motion picture.

Modern psychology has assimilated thoroughly the lesson taught by Thorndike and Dewey, that we learn by *doing* rather than by having actions demonstrated to us. While there are a few reports in the literature of "observational learning" among animals (3, 4, 11, 12) all the leading learning theories emphasize that the to-be-learned response must somehow be made to occur in the learning organism, not just in a model, before its association with a given set of conditions can be strengthened. Yet it is evident that viewers of movies and television dramas do learn much of the content of dramatic programs, at least to the extent of being able to recall subsequently many details of the characters' actions, and they sometimes incorporate material thus learned into their own overt behavior. There are instances, for example, in which criminals have employed techniques for crime taken directly from movie portrayals (1).

Presumably, a viewer's learning of the actions of movie characters is enhanced if he reproduces these actions himself, at least covertly, during the screen portrayal. Our problem in predicting which viewers will learn how much of what content then becomes one of understanding the conditions that affect a viewer's tendency to match his own actions, at least covertly, to those a screen character portrays.

We assume that when a viewer becomes absorbed in a dramatic production, he *identifies* himself, at least momentarily, with one or more of the characters. By identification we mean that the viewer, in fantasy, puts himself in the place of a character and momentarily feels that what is happening to that character is happening to himself. In this process, we assume that although he may reproduce very little of the gross motor behavior of his screen character (10) he does reproduce covertly[2] many elements of the behavior including the *emotions* he attributes to the character, so that when the character with whom he identifies himself is in a dangerous situation for example, the viewer feels fear, and when the character escapes danger, the viewer feels vicarious relief. We assume, too, that when a viewer puts himself in the place of a character, he orients toward the cues to which his character is responding. There may be misperceptions, of course, in which the viewer misjudges the nature of the cues to which the actor is responding, or misinterprets the actor's response, in which case the viewer learns something that is not an accurate representation of what the movie producer intended to portray.

While it is possible for a viewer to shift his identification from one character to another during a movie, most viewers probably make a choice of a character whom they consider to be the main protagonist of the story, and see the action mainly from his point of view. In many movies, the plot centers so fully upon one main character that all viewers may be expected to identify with him. But in other plots, there are two or more main characters sympathetically presented, and this makes it possible for viewers to differ in their choice of an "identificand." We assume that the nature of the content a viewer learns is in part determined by the choice of identificand, for the viewer presumably focuses his attention upon the stimuli that are relevant for his own character, and shares vicariously the experiences of that character rather than those of the people playing opposite him. Thus, if A assaults B on the screen, the viewers who are identified with A should feel anger and dominance, while those identified with B should feel fear, and these differential experiences at the time of viewing should be reflected in the nature of the

[1] This research was supported by a grant from the Laboratory of Social Relations, Harvard University.

[2] Covert practice has been shown to be fully as effective as overt practice in the learning of certain kinds of material from films. See Michael and Maccoby (6).

movie content that is learned. Of course, the matter of the ultimate incorporation of learned material into the viewer's own overt actions is a complex matter. This would depend not only upon the original learning of the material and perhaps its rehearsal in later spontaneous fantasy, but upon an appropriate situation arising in which the stored-up material can be evoked. Undoubtedly, much material assimilated from the mass media remains as an unused potential. A viewer may learn from a movie how to act as a groom at a wedding, but if he remains unmarried, the learning is never manifested in his behavior. Our concern in the remainder of this report is with the first step of learning from films; that is, with the subject's ability to recall film content verbally after a brief interval.

We expect, as noted, some individual differences in the choice of a movie character with whom to identify. We also assume that identification is a matter of degree, and that viewers differ in the "depth" of their identification with their chosen character at any particular moment. The deeper the identification with a character, the more should the viewer reproduce covertly, and therefore learn, the actions and feelings of that character.

In order to predict what a given viewer will learn, we must predict his choice of main protagonist, and the depth of his identification with that character. There are a number of presumably relevant factors. One undoubtedly is the consonance of the viewer's motives with the character's actions. If a screen character is doing what the viewer most wants to do, then, presumably, the viewer will be more likely to identify with him.

Another factor, we hypothesize, is that of similarity between the viewer and the screen character with respect to certain major social characteristics, such as sex, age, and race. The importance of similarity has been pointed out in connection with projective tests such as the TAT. Murray (9) has recommended that TAT pictures include a character of age and sex similar to that of the subject, to facilitate his projecting himself into a character. We believe that a similar process is involved in identifying with a movie character. While the fantasy sequence in a movie, unlike the TAT, is one over which the viewer has no control, the viewer *can* vary in the depth of his involvement in the story—the amount of attention he pays to it—and we are assuming that his interest will be more deeply engaged if the story has a central character who is similar to the viewer in some important and discernible aspect of the life roles he occupies.

There are instances, of course, in which a viewer's similarity to a character facilitates identification but in which the viewer's motives are more consonant with the actions of a different character. For example, the character in the story who is engaged in the activities the viewer would most like to perform may be of the opposite sex or a different social status. Under these conditions, we cannot make a prediction about the outcome.

Another factor which should affect the viewer's choice of a protagonist and the depth of his identification is the extent of the viewer's anxiety about the actions which the character portrays. Our expectation that a viewer should choose a protagonist similar to himself should hold if the actions being carried out are not only attractive but socially acceptable. But if the actions are simultaneously attractive and anxiety arousing, then the viewer would probably choose an identificand *dissimilar* to himself (14). The vicarious enjoyment of tabooed activity may be somewhat less if is enacted by a dissimilar person, but the anxiety it arouses would be reduced as well, and more sharply (cf. 7). We suspect that anxiety over the expression of aggression is one of the sources of children's preference for animal heroes and space men.

In the present studies, we have focused primarily upon the factor of *similarity*. That is, we have attempted to study whether viewers do in fact identify themselves with the movie character most like themselves, given a movie in which the viewer has some choice of protagonist, and we have attempted to determine whether differential identification produces predictable differences in the nature of the movie content learned. We have also given some attention to the matter of "consonance" between the motives of the viewer and actions of particular characters. We have *not* studied the third factor—displacement of a preferred agent of action due to anxiety—but have instead attempted to minimize the problem by choosing movies in which actions of a sympathetic character that might be potentially

anxiety provoking are justified in the plot so that guilt or fear will be minimal. Thus, when aggressive actions are carried out by a hero, they are justified by the villainy of a "bad man" and occur in a context of social approval.

STUDY I: SOCIAL CLASS

In studying the effects on identification of similarity between the viewer and a movie character, the first variable with which we worked was that of social class. The movie employed had two leading characters: a lower class boy and an upper-middle class boy. For this movie, our hypotheses were that:

1. Lower class children tend to identify with the lower class character and middle class children with the middle class character.

2. Viewers who identify with a character remember more of that character's words and actions than viewers who do not identify with that character.

3. Since both leading characters are boys, the predictions about identification apply more to boy viewers than to girl viewers.

Our hypotheses deal with the simple choice of a protagonist, not with the *degree* of involvement with that character once the choice has been made. However, consideration of degree would lead us to the parallel predictions that the more similar the viewer is to the movie character the more deeply he identifies with that character, and therefore the more his learning is specific to the content that is relevant to his chosen character.

We expected that our hypotheses might not apply uniformly throughout the movie. Specifically, the differences between the two groups of viewers (those identifying with the two different main characters) might be maximized either when both characters were on the screen at the same time, or when only one of them appeared on the screen. There were reasons, to be discussed, for either prediction. We therefore planned to use this study not only to test the specific hypotheses listed above, but to explore the group differences in selected scenes and for selected kinds of content.

Method

The Movie. Two 20-minute episodes of a serial movie entitled *Junior G-Men* were used in the study; it was a story about adolescent spy chasers made about 1940. Its two main characters were an upper-middle class boy and a lower class boy, both of whom were about 16 years of age. Harry Trent had his own car, lived in the best part of town, and could pilot an airplane and operate a radio transmitter. He was the leader of a group of boys who called themselves the Junior G-Men, and who were well dressed, polite, and had an elaborate club house. Billy Barton was the leader of a gang of "dead-end kids." They had an improvised club house in a junk yard and were ill-dressed and aggressive. Both Billy and Harry were admired fighters, but Harry won fights because he was a skillful boxer and Billy because he was a "slugger."

Subjects. The subjects (*Ss*) for this study were ten classes of seventh-grade students from public schools in the greater Boston area. The classes included 269 students—139 boys and 130 girls. Two or more classes were taken from each of four junior high schools serving communities that differed widely in socioeconomic level.

Each child in the study was asked to record his father's occupation at the end of his test blank, and the entries were coded on the Warner seven-point scale of occupational status (13). The sample included 94 children whose fathers' occupations fell at the first 3 steps of the Warner scale, 42 at step 4, and 126 at steps 5–7.

Procedure. The movie was shown to pairs of classes during a regular class period. The experimenter (*E*) told the children that they were about to see an ordinary entertainment film. The *E* said that he wanted to find out something about what kinds of things children notice and enjoy in movies, and that he would ask them a few questions about their reactions to the movie later on. The two reels shown were not consecutive, and while reels were being changed, the *E* described the intervening developments in the plot. When the two reels had been shown, the *E* said that he would come back in a week to ask some questions.

One week later, a multiple-choice questionnaire on the movie was administered. The items included questions on how well the *S* liked the movie, and several questions which were intended to reveal identification with particular characters, as well as test items on recall of the content of the movie. The *Ss* were told to make the best guess they could when they were not sure of an answer, and very few children omitted items in filling out the questionnaire.

Results

The first question of interest is whether children from lower-class families more often identified with the lower class hero than did the children with middle-class backgrounds. As we defined the term, identification meant the viewer's putting himself in the place of a given character as he experienced the story. Clearly, it is difficult to obtain a direct measure of this process. There was an attempt made to get an indirect measure by asking several questions concerning the *S*'s reactions to the main characters. We reasoned that a child would not

be likely to "put himself in the place of" a character whom he did not like and admire, since to do so would be to share disapproved characteristics. We also felt that the process of a viewer's imagining himself in a particular role is facilitated if the character playing the role behaves as the viewer himself would behave in a similar situation. We therefore asked three questions for our indirect measure of identification:

1. Which one of the two main characters did you like best?
2. Which of the two main characters would you like to be like?
3. Which of the two main characters do you feel is most like you?

The answers to these three items were combined into an identification index.[3] In the following discussion, a child is considered to be a "Billy" identifier if he chose Billy (the lower class boy) for two out of three of the above items. He is considered a "Harry" identifier if he chose Harry (the middle class boy) for two out of three of the items.

As may be seen from Table 1, the upper-middle class boys tended to choose Harry as an identificand and the lower SES boys tended to choose Billy. Among girls, a majority chose Harry, and this was true for girls at each social class level. Each of the items in the identification index contributed to the findings in Table 1. That is, each was positively related to social class among boys; the question "which character is most like you" contributed most, but the index correlated more highly with social class than did any individual item taken alone.

In Table 1, our measure of social class is a rating based on the occupation of the child's father. We asked each child not only about his father's occupation but about the kind of occupation the child himself would like to have when he was grown up. For girls this proved to be a difficult question, since a number wanted to be housewives, and those who did mention the kinds of jobs they would like to have tended to make choices such as secretary, nurse, teacher, or air-line stewardess which provided little range in aspired social class. For the following discussion of aspired social class, therefore, we have considered boys only.

[3] The three items had low positive intercorrelations, with coefficients of contingency ranging from .26 to .40.

TABLE 1

RELATION OF SOCIOECONOMIC STATUS TO CHOICE OF MOVIE CHARACTER*

	Socioeconomic Status (Warner Occupation Scale)	
	Upper Middle (1-3)	Lower Middle & Lower (4-7)
Boys		
Chose Billy	23%	56%
Chose Harry	75	40
Not ascertained	2	4
Total	100%	100%
Number of cases	(47)	(89)
$\chi^2 = 14.4$	$p = <.001$	
Girls		
Chose Billy	30%	33%
Chose Harry	64	61
Not ascertained	6	6
Total	100%	100%
Number of cases	(47)	(79)
$\chi^2 = .1$	$p = <.75$	

* Not included in this table are 7 cases whose SES was not ascertained.

It may be seen from Table 2 that, regardless of the social-class level of a boy's family, if he aspires to a relatively high level occupation, he tends to identify with Harry, the middle-class boy, in the movie. Boys who expect to have working-class occupations tend to identify with Billy, regardless of the occupation level of their fathers. Among boys with similar aspirations, there is only a little variation in choice of character that can be traced to current objective social-class membership. The conclusion is quite clear that a boy's choice of screen character is more closely related to the social-class level *to which he aspires* than to the level his family currently occupies.

Identification and the recall of movie content. Our hypothesis was that if a viewer is identified with a particular character in a movie, this fact should determine what events on the screen he perceives and remembers. Specifically, we predicted that viewers who identified with Billy should remember more of the things Billy said and did, and more of the things that happened to Billy, than should the Harry identifiers, and that conversely the Harry identifiers should remember more of the Harry-relevant material.

To test this prediction, we first designated the items in the multiple-choice test which

TABLE 2

RELATIONSHIP BETWEEN OCCUPATION TO WHICH BOY ASPIRES AND HIS CHOICE OF MOVIE CHARACTER*

Present and Aspired Social Class	Number Choosing Billy	Number Choosing Harry
Present SES High (1–3, Warner's Occupation Scale)		
Aspired occupation high	7	30
Aspired occupation low†	3	0
Present SES Low (4–7, Warner's Occupation Scale)		
Aspired occupation high	14	23
Aspired occupation low	30	10
$\chi^2 = 10.8$ $p = <.001$		

* Based on 117 boys for whom this information was available; 22 *S*s did not provide the information.

† Too few of the high SES *S*s aspired to low SES occupations to allow computation of χ^2.

could be classed as "Billy content" or "Harry content." The following exemplifies a Harry-relevant item: "In the plans for the boys to go to the Torchies' headquarters and try to capture the gang before they escaped, what was Harry supposed to do?" This question was followed by a list of possible answers. A Billy-relevant item was as follows: "What did Billy say when he turned around and faced the man who had followed him out on the docks?" In all, there were 27 items that could be classified as "Billy content," and 20 items for "Harry content." The test had 67 items, some of which were not classified into either of these content categories. Each child's test form was scored separately for the number correct on each of these two sets of items. Then each of these scores was divided by the child's "total correct" score for the entire test in order to eliminate the effects of reading ability, IQ, and other factors that might contribute to total score but were not relevant to our study.

Table 3 shows that Billy identifiers remembered slightly more Billy-relevant material than did the Harry identifiers. The differences, though significant for boys,[4] are not large. The comparisons on Harry-relevant content show that both boys and girls who identified with Harry remembered slightly more Harry-relevant content, but the differences are not significant.

[4] When the data are analyzed separately by each administration of the film, each administration being considered a replication, and the results combined, the Billy identifiers among the girls also remember more of the Billy content than do the Harry identifiers ($p < .05$).

TABLE 3

RELATIONSHIP OF CHOICE OF MOVIE CHARACTER TO RECALL OF MOVIE CONTENT

	Category of Content			
	Billy Content	Harry Content	Billy Aggression†	Billy Nonaggression†
Boys				
Mean Score* Billy Identifiers	44.0	28.0	19.4	24.8
Mean Score Harry Identifiers	42.1	28.7	18.0	24.3
t	2.25	1.00	2.48	.75
p‡	.02	n.s.	.01	n.s.
Girls				
Mean Score Billy Identifiers	44.6	28.0	18.3	25.7
Mean Score Harry Identifiers	43.3	28.2	17.2	26.3
t	1.10	—	1.36	.67
p‡	n.s.	n.s.	n.s.	n.s.

* Each *S*'s score is the proportion of all items answered correctly which were classified into the content category in question. Thus, among Billy identifiers, 44% of the correctly answered questions were Billy content items.

† The "Billy content" items were subdivided into aggression and nonaggression.

‡ All the p's are one-tailed unless otherwise stated.

A possible reason why these differences were so small is that the hypothesis tested does not take into account the attention a viewer devotes to his character's cues. We assumed that if a viewer is identified with a particular character he attends particularly to that character's actions and remarks. But it is probably also true that he attends to the stimulus properties of his character's environment. If, for example, the screen character is being approached by a dangerous animal, the viewer probably looks at the animal rather than the story hero. We could predict, then, that the viewer might notice and remember the cue properties of the hero's situation as well as (or better than) the actions and words of the hero himself. Thus, if the two main characters are on the screen simultaneously, and interacting with one another, each character's actions would be a cue for the other, and there should therefore be few differences in memory between the viewers who identify with each of the two characters.

To investigate this possibility, we subdivided

the test items according to whether the opposite character was present and interacting with the character asked about in the item. For boys, the Billy identifiers remembered significantly more of the Billy content when Harry was *not* on the screen with Billy. They remembered slightly, but not significantly, more Billy material when the two boys were present, but Billy's action did not involve talking to Harry. When Billy was talking to Harry, the Harry identifiers remembered the Billy content somewhat (not significantly) better.[6] For boys, then, the results fit the interpretation that when both characters are present and interacting, the nature of the remembered content is not a function of identification choice, since everything that happens is either the action of one's own character or a cue to which he responds.

We have shown so far that the different groups of viewers remembered much the same things from the movie but that the differences obtained favored the hypothesis that similarity of aspired-to class fosters identification with a given character, and that selecting one character for primary identification results in better learning of that character's actions and words. A major problem exists, however, in determining whether the differential learning is actually a result of identification. There is an alternative explanation that Billy identifiers remembered somewhat more Billy activity because Billy engaged in the kind of activities that were most interesting to the kind of boy who would choose Billy as an identificand, and that it was the interest in the activities per se, rather than identification with Billy, that produced the increment in learning. It is possible that Billy identifiers would have remembered more *Harry* material, if only the items about Harry had happened to concentrate upon his aggressive, motoric actions instead of his more restrained, verbal ones.

This alternative explanation is difficult to test, since we cannot guess what all the different sources of attraction for Billy-identifiers might be, nor do we have test items bearing upon them all. Partial tests were nevertheless possible with respect to aggressive versus nonaggressive content. Starting with the possibility that lower-class nonmobile children had a higher level of aggressive motivation than the other children (see 5) we considered whether the fact that Billy's activities in the movie were much more aggressive than Harry's could account for the greater learning of Billy content by Billy identifiers. We divided the test items into those dealing respectively with aggressive and with nonaggressive activity. Too few items on Harry aggression were available to make a test for Harry content, but we could compare the recall of Billy-aggressive content with Billy-nonaggressive content. As Table 3 shows, the superiority of Billy identifiers in recalling Billy content applied to the aggressive items only. Parenthetically, it may be of interest that boys recalled the aggressive content better, and girls the nonaggressive content, both differences being significant at the .01 level. Thus we find that the Billy identifiers did not show an advantage in recall of all kinds of Billy content, but only those selected aspects that were presumably consonant with their own motives. This finding would not be predicted from our view of identification, for when a viewer puts himself in a character's place, he should experience vicariously *everything* that character does and feels, regardless of whether each item of the character's behavior touches upon one of the viewer's major motives.

From our existing data we cannot tell whether identification has contributed to our results, or whether we are simply observing the effects of selective perception based upon need. If both identification and need-consonance were required for maximal recall of movie material, we should find that the Billy identifiers did not recall Harry aggression especially well, even though they are presumably set to select aggression, because the aggression is not carried out by their own preferred protagonist. In the absence of test data on Harry aggression, we cannot answer the question. More evidence bearing upon this problem is available from Study II.

STUDY II: SEX

In this study, we are concerned with the differences between boys and girls in their identification with movie characters and their

[6] For girls, there was a tendency for the Billy identifiers to show the greatest advantage over Harry identifiers in the content which involved Billy's talking to Harry.

memory of movie content. Continuing with our investigation of the hypothesis that viewers tend to identify with characters similar to themselves in major social categories, we sought a movie which would provide fairly comparable opportunities for identifying with either a male or female character. Our prediction was that viewers would choose to identify with the like-sexed character, and that the movie content they learned would be in part a function of this choice.

The specific hypotheses Study II was designed to test were as follows:

1. Boy viewers tend to identify with the boy character and girl viewers with the girl character.

2. Viewers who identify with a character remember more of the words and actions of that character than viewers identifying with the opposite character. This prediction applies only to words and actions of a given character that are *not* directed toward the opposite character.

3. Viewers who identify with a character remember more of the stimuli directed toward that character than those viewers who do not identify with that character. (Stimuli include words and actions of secondary characters directed toward the identified-with character, as well as other environmental cues to which he is responding.)

4. There is no difference between the viewers who identify with the two primary characters in the amount of material they remember when the two characters are interacting with each other.

5. Boys remember more of the aggressive content of the movie than do girls.

Method

The movie. The film was entitled *Tomboy*, a Hollywood feature made about 1940. The two primary characters were a boy and a girl, aged 14 or 15, who were pupils in a rural school. The story concerned a new girl in the community and a boy living with a cruel uncle, the developing interest of the boy and girl in each other, and the working out of the conflict between the boy and his uncle until the denouement with a happy ending.

Subjects. The *S*s in this study were 7th-grade pupils in three public schools in the Boston metropolitan area. The schools served quite varied communities: in one school, over half the fathers' occupations were classified in Warner's classifications 1, 2, and 3; in a second school, ⅔ of the fathers' occupations were in Warner's classifications 5, 6, and 7; in the third school, fathers' occupations were rather evenly distributed over the various classifications.

Administration. In each school, the movie was shown in an auditorium to all the 7th grade pupils. The *E* told the children that they were about to see an ordinary entertainment movie; he said that he wanted to find out something about what kinds of things children notice and enjoy in movies, and that he would ask them a few questions about their reactions to the movie later on. One week later the *E*s returned with questionnaires. The pupils remained in their classrooms and were tested one classroom at a time.

Measuring instruments. The measuring instruments were comprised of two parts: one designed to provide an indirect measure of identification; and the second designed to measure learning of the content of the movie. In addition, each *S* was asked to indicate his sex and his father's occupation.

As in Study I, we do not have a direct measure of the *S*'s "identification" with a movie character, but employed an indirect measure intended to reveal aspects of each role's attractiveness for fantasy role playing. Since we found in the first study that the question "which character did you like the best" was not as useful in predicting the recall of movie content as the other items in our identification index, we omitted this question, and used the following questions: (*a*) "Which character is most like you?"; (*b*) "Which character would you like to be in real life?"; (*c*) "Which part would you like to play in the movie?" The *S*s were given five choices—the two primary characters and three secondary characters—and were asked to indicate their first two choices to each question. Their answers were combined into an index, described below.

There are four kinds of content relevant to the hypotheses guiding this study: (*a*) behavior, including actions, speech, and feelings, of one primary character not directed toward the other primary character (designated "Alone" content); (*b*) behavior of one primary character directed toward the other primary character when they are in a scene together (designated "Interaction" content); (*c*) stimuli directed toward one of the primary characters but not by the other (designated "Stimulus" content); and (*d*) behavior of a primary character directed toward a secondary character in anger and/or with the purpose of injuring him either physically or psychologically (designated "Aggressive" content).

Three different procedures were used, on different groups of *S*s, to measure the learning of content:

The first was a multiple-choice "recognition" test, similar to that used in Study I. This test was given to 291 students. In Study I there were disappointingly small differences in recall scores between viewers identifying with the two main characters. This fact suggested that our measuring instrument was relatively insensitive, and in Study II we therefore tried out two additional kinds of measuring devices in an effort to obtain sharper differences. One was a "recall" test, in which the *S*s (47 in number) were required to write in their own answers to "open" questions. Another was a "reconstruction" test, given to 37 students, to be described more fully below.

Results and Discussion

Identification. An "identification index" was computed on the basis of the answers to the three identification questions; this index indicated the degree to which one of the primary characters was seen as a more attractive role than the other. In computing index scores, each first choice was given a value of two, and each second choice a value of one. The index was equal to the score for the primary boy character minus the score for the primary girl character. Index scores showed very little cross-sex identification, as may be seen in Table 4. We can therefore test our hypotheses in terms of simple sex differences. All the following comparisons are thus in terms of boys versus girls.

Learning. As is so often the case at this age with verbal materials, the girls performed at a somewhat higher level of skill on all three measures of learning of movie content. For example, on the recognition test the girls got an average of 39.5 items correct, as compared with an average of 38.0 for the boys; this difference is significant at the .03 level. To allow for this difference and within-sex differences in total score, all the raw scores for each category of content were converted into percentage scores on the basis of the individual *S*'s total number correct. The results for the three different types of measuring instruments are considered separately below.

Recognition. The recognition test was composed of 64 multiple-choice items about the content of the movie. There were four kinds of content relevant to our study, and some items in each of these content categories concerned the male leading character, some the female. The items were selected in terms of (and classified into) the following eight categories:

1. Boy Alone—boy character's actions and words when the girl was not present. 11 items. (Example: "How did Steve react when his Uncle Mat told him he couldn't go to school anymore?")

2. Boy Interaction—boy character's actions and words directed toward the girl. 7 items. (Example: "When Pat and her father came by to see Uncle Mat on Sunday afternoon, what did Steve say to Pat when he first got out to the car?")

3. Boy Stimulus—stimuli directed toward the boy character but not by the girl. 11 items. (Example: "Steve was talking to Pat and her father on Sunday morning and suddenly said, 'Gee, it's late. I've got to run.' What reminded him he had to go?")

4. Boy Aggression. 7 items. (Example: "How many times did Steve hit Harry?")

5. Girl Alone. 7 items. (Example: "When Pat was getting lunch ready on the picnic, how did she show that she missed Steve?")

6. Girl Interaction. 7 items. (Example: "When Pat and her father stopped by at Steve's house after the picnic, what did Pat say to Steve?")

7. Girl Stimulus. 7 items. (Example: "When Pat went out the back door on Sunday morning looking for her father, what did she see?")

8. Girl Aggression. 7 items. (Example: "What did Pat do while the tramp was holding her?")

We predicted that boys would remember more aggressive content from the film than girls. This proved to be true. (See Table 5-A.) Furthermore, girls remembered more Girl Alone content, which was predicted, and more Girl Interaction content, which was *not* predicted. There were no significant differences on the other categories of content, although all were in the predicted direction.

Recall. The instrument used in the recall testing was composed of items that asked a question about the content of the movie and provided a space for the *S*'s response. The items were selected from ones used in the recognition procedure, the stems being slightly reworded. There were five items from each of the eight categories of content. In scoring the recall questionnaire, two points were given for each essentially correct item, one point given for a partially correct item, and 0 points for each incorrect item. Points were added to

TABLE 4

IDENTIFICATION WITH BOY AND GIRL MOVIE CHARACTERS BY SEX OF VIEWER

	Choose Boy Character						No Choice	Choose Girl Character					
Identification Index Score	6	5	4	3	2	1	0	−1	−2	−3	−4	−5	−6
Number of Boys	44	34	46	17	20	12	13	8	3	2	1	0	0
Number of Girls	0	1	1	1	0	6	7	7	13	26	23	37	53

TABLE 5
SEX DIFFERENCES IN MEMORY OF MOVIE CONTENT

	Content Categories							
	Boy Alone	Boy Stimulus	Boy Aggression	Boy Interaction	Girl Alone	Girl Stimulus	Girl Aggression	Girl Interaction
A: Recognition Test*								
Mean Boys' per cent score (N = 153)	18.0	16.6	8.6	10.7	11.5	13.7	9.7	11.3
Mean Girls' per cent score (N = 138)	17.9	16.4	7.4	10.7	12.3	14.2	9.5	12.1
t	.33	.53	3.50	.17	2.28	1.44	.56	2.12
p†	n.s.	n.s.	.001	n.s.	.01	.10	n.s.	.05‡
B: Recall Test*								
Mean Boys' per cent score (N = 27)	16.00	10.1	19.3	6.9	12.7	12.7	10.6	13.2
Mean Girls' per cent score (N = 20)	14.50	8.2	13.6	8.1	10.9	15.1	10.2	19.5
t	.78	1.36	3.98	.81	1.11	1.19	.25	3.15
p†	n.s.	.10	.001	n.s.	n.s.	.12	n.s.	.005‡
C: Reconstruction Test: Unit Classification*								
Mean Boys' per cent score (N = 20)	31.4	19.6	—	12.7	19.0	8.5	—	4.2
Mean Girls' per cent score (N = 17)	26.2	17.5	—	12.3	25.1	10.5	—	3.5
t	2.32	1.17		.20	3.38	1.70		.64
p†	<.02	<.13		n.s.	<.001	<.05		n.s.

* Each S's score is the number of items correct in a content category divided by the total number correct on the entire test.
† All p's are one-tailed tests unless otherwise indicated.
‡ P is two-tailed: no direction was predicted.

provide a point score for each category of content and a total point score.

On the recall test, boys remembered more Boy Aggression content than did girls, and girls remembered more Girl Interaction content (see Table 5-B). Differences here are much larger than for the previously discussed recognition test. The differences obtained for Boy Aggression and Girl Interaction scores were at the same or higher level of statistical significance even though the test was shorter and only about ⅓ the number of Ss used.

Reconstruction. The reconstruction testing procedure involved the use of still photographic prints of scenes from the movie. The Ss were to respond to each stimulus picture by writing a paragraph or two telling what was happening at the moment of the picture, what had happened just before the picture, and what happened just after the picture. They were individually given nine 3 x 4½ inch photographic prints, with four minutes to write about each. Three pictures depicted the primary boy character alone or with secondary characters; three pictures depicted the primary girl character alone or with secondary characters, and three pictures depicted the two primary characters, boy and girl, together.[6]

[6] Unfortunately, in printing, the pictures were reversed from right to left. The Ss did not report being confused about the scenes on this account, however.

The scoring of the resulting protocols was in terms of a "correct and relevant subject-predicate unit"; that is, each subject-predicate unit that actually occurred in the movie in the specific scene from which the picture was taken was counted. In addition, a score of one unit was given for the mention of material—no matter what the quantity in terms of subject-predicate units—that did not occur in the specific scene in question but was relevant to it either as a direct antecedent of the scene or a direct consequence of it. Material that actually occurred in the movie but was not directly relevant to the scene in question was not counted, nor was material that did not occur in the movie.

Each unit was classified if possible in one of the eight previously defined relevant categories. Less than two per cent of the content of the protocols could not be so placed. Aggressive material made up such a small proportion of the content of these protocols (about 3%) that it is not considered in the comparisons.

Half of the protocols were scored by two independent scorers. Interscorer reliabilities were computed by means of rank-order correlations. The reliability of the total scores was .90, while the reliability of the scores on the nine individual pictures ranged from .81 to .95 with a median of .86, and that of the scores for the six subcategories of content ranged from

.65 to .94 with a median of .80. Score-rescore reliabilities of a single scorer were somewhat higher.

The results obtained through this procedure may be analyzed in two ways: first, in terms of the individual unit classification, and secondly, in terms of the picture classification. In the first instance, boys and girls are compared on the basis of the score for each category of scorable units (Boy Alone, etc.), without regard for the classification of the picture for which the units were written. In the second instance, boys and girls are compared on the basis of the scores on each type of picture: Boy Alone, Girl Alone, and Interaction, without regard for the classification of each particular unit scored for that picture.

In the first analysis, boys obtained significantly higher scores on Boy Alone content; girls obtained significantly higher scores on Girl Alone and Girl Stimulus content; and there were no significant differences between boys and girls on Boy Stimulus, Boy Interaction, and Girl Interaction content (see Table 5-C). It should be noted that the differences obtained with the use of the reconstruction test are larger than those obtained on the recognition test.

In the second analysis, boys obtained statistically significant higher scores on the Boy Alone pictures; girls obtained significantly higher scores on the Girl Alone pictures; and there was no significant difference between the boys and girls on the Interactive pictures. (See Table 6.)

Combined results. Since the content categories for the three testing procedures were defined in essentially similar ways and the three procedures were used on independent sets of *Ss*, it is permissible to combine the probabilities obtained on the three procedures by the Stouffer technique (8). The results are shown in Table 7.

TABLE 6

SEX DIFFERENCES IN MEMORY FOR MOVIE CONTENT RECONSTRUCTION TEST: PICTURE CLASSIFICATION

	Classification of Picture		
	Boy Alone	Girl Alone	Interactive
Mean Boy per cent score ($N = 20$)	35.3	27.0	38.3
Mean Girl per cent score ($N = 17$)	30.4	33.9	35.6
t	2.43	3.52	1.21
p*	$< .01$	$< .001$	$< .25$

* P's are one-tailed tests except for the "Interactive" which is two-tailed.

Some of our original hypotheses are supported; others are not. We predicted that boys would remember more of the Boy Alone and Boy Stimulus content, and this proved to be the case. We also predicted that boys would remember more aggressive content, but we found that they remembered the aggressive content better than the girls only if the boy hero was the agent of aggression. As far as the girls are concerned, they remembered more Girl Alone and Girl Stimulus material, as predicted, but their greater memory for "girl" content did not extend to aggression, even when the girl was the agent of aggression. Furthermore, we predicted that there would be no difference between the sexes in their recollection of content in which the boy and girl were interacting, on the grounds that any action on the screen would constitute either an action of one's own character or a relevant cue for his behavior. We found, however, that girls *did* remember, better than boys, the words and actions of the girl character when they occurred in a context of interaction with the boy.

These findings suggest certain revisions of our hypotheses, still further in the direction

TABLE 7

COMBINED PROBABILITIES OF SEX DIFFERENCES OBTAINED IN THREE PROCEDURES OF STUDY II

	Content Categories							
	Boy Alone	Boy Stimulus	Boy Aggression	Boy Interaction	Girl Alone	Girl Stimulus	Girl Aggression	Girl Interaction
Recognition Test Z ($N = 47$)	0.32	0.52	3.50	0.16	2.27	1.43	0.56	2.12
Recall Test Z ($N = 47$)	0.77	1.34	3.68	−0.80	−1.09	1.17	0.25	2.90
Reconstruction Test Z ($N = 37$)	2.22	1.15	—	0.20	3.08	1.65	—	−0.63
Combined Z	1.91	1.74	5.10	−0.25	2.46	2.45	0.57	2.54
Combined p (one-tail)	$< .03$	$< .05$	$< .001$	$< .50$	$< .01$	$< .01$	$< .30$	$< .01$

indicated in Study I. Memory for movie content appears to be influenced *both* by identification with a character in the movie and by the "need relevance" of a particular kind of content for the viewer. In cases where both identification and relevance are acting in the same direction, one should expect the greatest differences between boys and girls; where identification and relevance are acting in different directions, one should expect a minimum of difference between boys and girls; and where only identification is affecting the situation, the differences should be intermediate. If we assume that aggressive material is particularly relevant for boys, and boy-girl interactive material for girls,[7] the differences would be almost exactly of the order we find in Table 7, with Boy Aggression and Girl Interaction showing substantial differences, and Girl Aggression and Boy Interaction showing little or no differences.

SUMMARY

Our hypothesis was that the amount and kind of material a viewer learns from a movie is in part a function of his choice of a character in the movie with whom to identify himself. We expected that a viewer would tend to identify with a character similar to himself in major role characteristics (e.g., sex, age, or status) whenever the movie provided more than one sympathetically presented character. And we reasoned that the viewer, having chosen a character as the primary protagonist through whose eyes to experience the action of the movie, would remember more of that character's actions and words, and more of the stimuli to which that character was responding, than would the viewer who was identified with a different character.

In two studies we showed movies to 25 class rooms of seventh-grade children, and tested them a week later on their knowledge of movie content.

Using an indirect measure of identification, we found that viewers did indeed identify themselves with the like-sexed leading character, in viewing a movie which included both a strong male and strong female lead. With respect to similarity of social class, however, viewers were more likely to choose the protagonist whose social class corresponded with the viewer's *aspired* social class, rather than his current objective status.

As far as memory for movie content was concerned, we found that both the choice of a character with whom to identify and the "need relevance" of that character's actions were important. That is, viewers do tend to remember somewhat better the actions and words of the character with whom they identify, and the stimuli relevant for that character's actions, but the advantage in learning does not apply equally to *all* the actions of the viewer's chosen character. For boy viewers, aggressive content appears to be particularly relevant, for they remember aggressive content better than girls, *provided* its agent is the boy hero. For girls, boy-girl interactive content appears most relevant, and they remember this content better than boys whenever the girl heroine is the agent of the action.

[7] Our assumption about aggressive material being more relevant for boys, and romantic material for girls, is consonant with the findings of Dysinger & Ruckmik (2). These workers took PGR readings of children watching movies, and found a higher level of emotional response among boys to danger and conflict scenes, and among girls, to love and erotic scenes.

REFERENCES

1. BLUMER, H., & HAUSER, P. M. *Movies, delinquency and crime*. New York: Macmillan, 1933.
2. DYSINGER, W. S., & RUCKMIK, C. A. *The emotional responses of children to the motion picture situation*. New York: Macmillan, 1933.
3. HAYES, K. J., & HAYES, C. Imitation in a home-raised chimpanzee. *J. comp. physiol. Psychol.*, 1952, **45**, 450–459.
4. HERBERT, J. J., & HARSH, C. M. Observational learning by cats. *J. comp. Psychol.*, 1944, **17**, 81–95.
5. MCKEE, J. P., & LEADER, FLORENCE B. The relationship of socio-economic status and aggression to the competitive behavior of preschool children. *Child Developm.*, 1955, **26**, 136–142.
6. MICHAEL, D. N., & MACCOBY, N. Factors influencing verbal learning from films under varying conditions of audience participation. *J. exp. Psychol.*, 1953, **46**, 411–418.
7. MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.
8. MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge: Addison-Wesley, 1954.

9. MURRAY, H. A. *Explorations in personality*. New York: Oxford Univer. Press, 1938.
10. ROSE, N., FAISON, E., & PODELL, J. Effects of rest pauses during a training film upon audience responses. *Amer. Psychologist*, 1953, **8**, 424. (Abstract)
11. WARDEN, C. J., FIELD, H. A., & KOCH, A. M. Imitative behavior in the Cebus and Rhesus monkeys. *J. genet. Psychol.*, 1940, **56**, 311–322.
12. WARDEN, C. J., & JACKSON, T. A. Imitative behavior in the Rhesus monkey. *J. genet. Psychol.*, 1935, **46**, 103–125.
13. WARNER, W. L., MEEKER, M., & EELLS, K. *Social class in America*. Chicago: Science Research Associates, 1949.
14. WRIGHT, G. O. Projection and displacement: A cross-cultural study of folk-tale aggression. *J. abnorm. soc. Psychol.*, 1954, **49**, 523–528.

Received September 17, 1956.

BODY-IMAGE BOUNDARIES AND ADJUSTMENT TO POLIOMYELITIS

KENNETH E. WARE

Baylor University College of Medicine

Southwestern Poliomyelitis Respiratory Center[1]

SEYMOUR FISHER AND SIDNEY CLEVELAND

Veterans Administration Hospital, Houston

It has been a general observation that among patients afflicted with physically crippling illnesses or injuries, the psychological reaction to the trauma seems to bear little relationship to the extent or degree of the disability. Some patients with minimal physical incapacitation demonstrate a catastrophic reaction, a feeling of complete devastation, while others with extensive physical involvement are able to accept the prospects of permanent severe crippling and adjust to their handicap without evidence of marked personality disruption. Perhaps this difference in psychological reaction to a change in body structure may be accounted for, at least in part, by differences in the way individuals conceive of their own bodies. The concept of body image as a system of ideas and feelings which the individual has about his physical structure has had frequent application in the psychological literature on poliomyelitis and other disabling conditions. It has been pointed out (1, 5, 6, 9, 10) that disablement imposes upon the individual the necessity for a more or less drastic modification of his body image, an abrupt change in his perception of himself as a physical being. Difficulties in adjustment are often conceptualized as difficulties in incorporating the atrophied body, the facial disfigurement, or paralyzed limb into the basic, pre-existing body image. The ease or difficulty with which this integration takes place is an index of the success which one has with one aspect of the problem of readjustment.

Although body image is a concept often used in discussions of the psychological implications of physical handicap, there have been few attempts to derive systematic measurements relating to this phenomenon. Fisher and Cleveland (2, 3, 4), however, have described a method of obtaining measurements on one dimension of body image. This dimension, called the body-image boundary, has to do with the degree to which the individual conceives of his body boundaries as definite and firm versus indefinite and weak. Its measurement is expressed in terms of what is called the Barrier score or the total number of a particular type of responses (called Barrier responses) produced by the individual on the Rorschach test. In addition to their published findings on normals and various psychosomatic symptom groups, Fisher and Cleveland have determined in unreported exploratory studies that degree of boundary definiteness is significantly linked with ability to tolerate stress in normal subjects and with facility of readjustment to body mutilation in individuals who have suffered limb amputation.

The concept of body-image boundary should have especially marked implications in a disease such as poliomyelitis in view of the fact that probably no other disease carries with it the actuality or threat of more severe physical incapacitation. It is conceivable that successful adaptation to such a disability may be positively related to the strength or firmness of the body-image boundary. The individual who already has a body image with a firm, definite, and protective boundary should be better able to withstand the threat incurred from physical insult, to incorporate the damage, and to reestablish the body-image boundary.

Within this framework, a study was designed to test the hypothesis that patients who achieve the better psychological adjustment to the physical disability imposed by poliomyelitis show, through a greater number of Barrier responses, evidence of a stronger, more definite body-image boundary than patients with a less satisfactory adjustment. Further, it was argued that the Barrier score, rather than reflecting the degree of the individual's physical incapacitation or the duration of his illness at the time of testing, is a

[1] Southwestern Poliomyelitis Respiratory Center, is aided by an annual grant from the National Foundation for Infantile Paralysis, Inc.

function of a more pervasive body-image factor in the personality which may dispose the individual toward a successful or unsuccessful adaptation to physical handicap.

METHOD

The subjects (*Ss*) were 56 patients hospitalized in a poliomyelitis respiratory center which cares for patients in both the acute and postacute stages of illness. The group included all types of poliomyelitis cases: spinal, bulbar, and bulbo-spinal, both with and without respiratory muscle paralysis. It included cases with widely varying physical involvement, ranging from relatively mild to severe with a predominance of severe cases. It was regarded as representative of the usual patient population of the respiratory center. The group included 29 males and 27 females. The age range was from 15 to 43 years with a median of 26; education varied from 3 to 20 years with a median of 13; and Wechsler Verbal IQs (available on all but 4 patients) ranged from 72 to 145 with a median value of 114.[2]

In order to obtain the sample, files were examined on all patients, both current and discharged. Sixty *Ss* were found whose Rorschach protocols contained the minimum number of responses necessary for scoring the body-image variable. These patients were then evaluated in terms of adjustment to illness by procedures which will be described later. In the process of this evaluation, four patients were eliminated because of disagreement between judges as to adequacy of adjustment, leaving a total of 56 *Ss*.

The Rorschach test had been administered individually to each of these patients for general clinical assessment purposes prior to the inception of this study. In all cases the test had been given by an examiner who at the time had no knowledge of the method of determining body-image values from the Rorschach. Since the test had been given more or less routinely to adult patients, there was no indication of a selective factor in the administration.

Twenty-eight of the 56 *Ss* had been tested by the senior author and at the time of the study were either still in the hospital or had only recently been discharged. The remaining 28 *Ss* had been tested and subsequently discharged prior to the senior writer's employment at the respiratory center. These two groups, which were treated separately as independent samples, are referred to as Groups I and II, respectively. Adjustment ratings were made in somewhat different fashion for the two groups and involved different sets of judges. It was of interest to determine whether a relationship between body-image data and adaptation to disability occurred in both samples. If so, it would suggest a degree of stability in the relationship.

Groups I and II were each subdivided into two categories relating to the adequacy of adaptation to illness. The method used was that of gross clinical evaluation. In the case of patients in Group I, three staff physicians who knew the patients well and a psychologist (the present senior author), made independent gross evaluations of the 32 *Ss* on whom Rorschach records of sufficient length were available. These evaluations were made prior to inspection of the Rorschach protocols with reference to the body-image variable. Judges were asked to divide this group of patients into two categories of approximately equal number, placing those who had shown the most satisfactory adaptation to their illness in one group, and those with the least satisfactory adaptation in another. No attempt was made to define for the judges the criteria for satisfactory or unsatisfactory adjustment. They were simply told to use their global clinical impression of the patient's behavior during his period of hospitalization in allocating him to one or the other of the two categories. In 28 out of the 32 cases, at least three out of the four judges agreed as to placement of the patient. These 28 individuals composed Group I. The "satisfactory" and "unsatisfactory" categories thus identified included 15 and 13 patients, respectively.

Because the 28 patients in Group II had been discharged for some time, it was felt that staff members would not remember them well enough to make an evaluation of their adjustment while in the hospital. However, an index of adjustment was already available in the records of these patients. Soon after the patient's discharge, the social worker at the Center had rated on a 5-point scale the adequacy of the patient's overall psychological adjustment to his illness during the course of his hospitalization. This evaluation in a sense represented the pooled judgments of the entire staff. It was essentially an integration by the social worker of her own impressions and the opinions expressed by various members of the staff during a number of conferences on the patient and in the course of informal contacts between the social worker and other staff members while the patient was in the respiratory center. This entry, known as "Patient's Response to Disease (PRD)," is one dimension of the "Social Service Profile,"[3] which has been in use at the Center for a number of years. It is completed on each patient by the social worker as a routine procedure. For purposes of illustration, the definitions of the two extremes of the PRD scale are shown as they appear in the Profile:

> *PRD O.* Degree of anxiety and apprehension present does not seem out of proportion to the neurological involvement; cooperates to the extent of his ability. His responses to illness and hospitalization do not interfere with his acceptance of treatment. Can utilize his residual abilities rather than withdrawing and becoming completely dependent on

[2] This high median intelligence level should not be taken as an indication of selective sampling. The Wechsler-Bellevue Verbal Scale has been given to more than a hundred adult patients at the respiratory center by various examiners during the past three years, and the median IQ for these data corresponds very closely with the median IQ of 114 reported for the *Ss* used in this study. Apparently adult poliomyelitis patients, at least those treated in this center, are as a group above average in intelligence.

[3] The "Social Service Profile" was designed by Maurine B. Mitchell, Director of Medical Social Work, Southwestern Poliomyelitis Respiratory Center and Wolff Home Poliomyelitis Rehabilitation Unit.

others; is considerate of the rights and privileges of others (within his physical limitations).

PRD 4. Physiologically unjustified amount of extreme anxiety; insecurity and apprehension prevent cooperation; may result in patient becoming either fearful, belligerent, hostile, or rebellious; fails in making acceptable adjustment to hospitalization and separation from family; complete obstruction of treatment program.

For the patients in Group II, dichotomizing the distribution of PRD scores into categories of 0–1 and 2–4 yielded two groups of 14 patients each, which were designated as "satisfactory" and "unsatisfactory" adjustment groups.

Descriptive data on the total sample and on the four subgroups are presented in Table 1.

The method used for scoring body-image boundary definiteness from the Rorschach has been described in detail elsewhere (2). The basic assumption underlying the scoring procedure is that the definiteness of the individual's body-image boundaries is reflected in the manner in which he describes the peripheries of his ink blot percepts. Definite boundaries are linked with descriptions which emphasize the protective, decorative, covering, containing, and concealing attributes of percept peripheries. The following are examples of Rorschach responses which exemplify the attributing of such special qualities to the periphery: "cave with rocky wall," "mummy all wrapped up," "man in suit of armor," "fort with high walls," "woman with fancy decorative dress." The number of responses of this sort given by an individual in his Rorschach record is taken as an index or score of boundary definiteness and is referred to as the Barrier score. It has been shown that the Barrier score can be determined by scorers with a high degree of objectivity (2).

Because the Barrier score is probably correlated with the total number of Rorschach responses produced, it was necessary to control variations in response total as much as possible. This was done by using only records with at least 15 responses and by reducing to 25 responses all records with more than 25. In the case of records with R greater than 25, this procedure involved using only the first three responses on each of the first five Rorschach cards and the first two responses on each of the last five cards. When there was less than the set quota of responses on any given card, it was compensated for by taking an extra response on the next card in the sequence for which there was more than the set quota.

After the records had been scored for Barrier responses, using this method of controlling R, the group of 56 *S*s was dichotomized at the median of the Barrier score distribution to yield a high- and a low-Barrier group. It was found that these two groups each had a response total of 25. Although the range of responses was 15–25, actually only a few cases in each group fell below 25, so that the median in both cases corresponded to the upper limit of possible response totals. Since the medians were the same, it may be concluded that there was no significant difference between high- and low-Barrier *S*s with respect to R.

Intelligence was a second factor to be considered in addition to R as possibly affecting the production of Barrier scores. The median IQ for the high-Barrier group was 114 and for the low-Barrier group 113.5. Accordingly, it was assumed that intelligence did not contribute to the difference in Barrier scores between the two groups.

RESULTS

Testing the hypothesis, which asserted a relationship between body-image boundary and adjustment to illness, required a comparison of "satisfactory" and "unsatisfactory" groups with respect to the frequency of Barrier scores produced. As shown in Table 2, the median Barrier score for patients in subgroups characterized by the more satisfactory adjustment seemed to exceed that of patients who made less satisfactory adjustments. Within each of the two samples (Groups I and II), chi square was used as the test of significance in comparing "satisfactory" and "unsatisfactory" patient groups with regard to the number of *S*s above and below the median of the Barrier score distribution. The Barrier score frequency

TABLE 1

CHARACTERISTICS OF SUBJECTS CLASSIFIED AS HAVING "SATISFACTORY" AND "UNSATISFACTORY" ADJUSTMENT AND OF THE TOTAL GROUP

Group and Adjustment Category	*N*	Sex		Age		Education		W-B Verbal IQ		Months Ill at Testing*	
		M	F	Mdn	Range	Mdn	Range	Mdn	Range	Mdn	Range
Group I											
Satisfactory	15	6	9	26	20–34	12	9–17	113	76–129	3	1–14
Unsatisfactory	13	6	7	26	19–43	13	9–17	113	72–132	8	1–17
Group II											
Satisfactory	14	10	4	25	18–38	15	3–17	115	80–133	7	1–14
Unsatisfactory	14	7	7	27	15–33	12	6–20	119	81–145	12	6–18
Total group	56	29	27	26	15–43	13	3–20	114	72–145	7	1–18

* By Rorschach test.

TABLE 2

MEDIAN AND RANGE OF BARRIER RESPONSES FOR SUBJECTS CLASSIFIED AS HAVING "SATISFACTORY" AND "UNSATISFACTORY" ADJUSTMENT AND FOR THE TOTAL GROUP

Group and Category	Median	Range
Group I		
Satisfactory	4	1–7
Unsatisfactory	2	1–7
Group II		
Satisfactory	4.5	0–8
Unsatisfactory	2.5	1–5
Total group	3	0–8

was distributed in categories of 0–3 and 4 or more responses. Correction for continuity was used in both analyses. Results of these procedures are shown in Table 3.

Examination of Table 3 reveals that the incidence of Barrier scores in both Groups I and II was in the predicted direction with patients in the "satisfactory" categories showing the greater frequency. In the case of Group I, the difference in Barrier score frequency between the "satisfactory" and "unsatisfactory" categories was significant at the .11 level and for Group II at the .06 level of confidence. Since the assumption of independence of experimental outcomes was satisfied, the two probabilities were combined to yield a composite chi square (8, pp. 46–47). The probability associated with this chi square would be the aggregate probability of obtaining the given set of results from two independent tests of the relationship between the adjustment and body-image variables. The composite chi square was found to be 10.06. With four degrees of freedom (twice the number of probabilities involved), this value for chi square falls between the five and the two per cent levels of significance. Thus, although results of the two independent analyses did not prove to be strictly significant, the collective results support the hypothesis. In other words, it would appear that people who show evidence of definite body-image boundaries, in comparison with those having indefinite boundaries, tend to adjust significantly better to the experience of physical disability resulting from poliomyelitis.

The question remains of whether the Barrier score reflects a relatively well established body-image factor in the personality, or is simply

TABLE 3

COMPARISON OF PATIENTS CLASSIFIED AS HAVING "SATISFACTORY" AND "UNSATISFACTORY" ADJUSTMENT ON FREQUENCY OF BARRIER RESPONSES

Group	Satisfactory		Unsatisfactory		χ^2	p
	Above Mdn	Below Mdn	Above Mdn	Below Mdn		
I	9	6	3	10	2.58	.11
II	9	5	3	11	3.65	.06

a function of the degree of physical incapacitation or duration of illness at the time of testing. Three steps were taken in examining this problem. The chief physician was asked to pick from the total group the 15 patients whom he considered to be the most incapacitated from the physical standpoint, and the 15 patients with the least physical involvement. This procedure was used because accurate ratings on degree of physical disability could be made only for extremes. Chi square was used as before in comparing the two groups with respect to the number of individuals with Barrier response totals above and below the median of three. The results showed no difference in Barrier scores for groups characterized by different degrees of physical involvement.

As a second step, the relationship of Barrier scores to duration of illness at time of testing was studied in the total group by means of chi square. It was found that length of illness at time of testing for high-Barrier *S*s did not differ significantly from the duration of illness at time tested for low-Barrier *S*s. Median durations for the high- and low-Barrier groups were 7 and 8 months, respectively. It should also be noted that Fisher and Cleveland, in a previous study (2), found no relationship between Barrier score and duration of illness in a group with rheumatoid arthritis and a group with neurodermatoses.

These observations are consistent with the argument advanced here, but it is evident from Table 1 that both of the groups rated "unsatisfactory" had been ill for a considerably longer period of time when tested than those rated "satisfactory" on adjustment. This raised the question of whether the significant relationship between Barrier score and adjustment might be contaminated by the variable of duration of illness. In order to control the

possible influence of this variable, a matched pair technique was utilized as a third step in the analysis.

It was possible to select from the total sample a group of 15 "satisfactory" *Ss*, each of whom was closely matched (within one month) with respect to duration of illness by a patient among the "unsatisfactory" *Ss*. The median duration at time of testing was 7.5 months, and the range was 1 to 14 months for each of the two groups thus obtained. Chi square, corrected for continuity, was used to compare the two groups with regard to the number of *Ss* above and below the Barrier score median. With length of illness at time tested held constant, high-Barrier individuals still fell more frequently in the "satisfactory" category (.06 level of confidence). This result did not quite reach the conventional criterion for statistical significance, but it was regarded as a fair indication that the Barrier score is not (or at least not primarily) a function of how long the patient had been ill when he was tested.

It might be noted, incidentally, in speculating as to why the more poorly adjusted patient was tested at a later date in his hospitalization than someone with a better adaptation to his illness, that the most plausible explanation is that the patient who made a poor adjustment was simply a more disagreeable and more difficult individual to work with and was likely to be resistant to testing. There was probably a tendency to delay and procrastinate in seeing this patient, especially since it was known in most cases that he would be available for testing over a period of several months of hospitalization.

DISCUSSION

The results confirm the major hypothesis involved in this study and also furnish evidence that would seem to make the argument advanced for the nature of the Barrier score a tenable point of view. The correlation found to exist between the body-image variable and adjustment seems to indicate that the extent to which the individual conceives of his body exterior as having a defensive, shielding, barrier-like quality plays some part in the total constellation of forces affecting the adequacy of the individual's adaptation to physical crippling. Apparently, the individual whose body image includes these qualities is better able to cope with the problems imposed by his handicap than is the person whose body concept lacks these characteristics.

It is readily evident, in the light of the considerable degree of overlap in Barrier score frequently between "satisfactory" and "unsatisfactory" groups, that the body-image variable, at least as presently defined, possesses little diagnostic or predictive efficiency. Nevertheless, the obtained relationships suggest that the concepts involved are theoretically fruitful and merit further inquiry.

In general, the Barrier score seems to reflect some kind of a body-image factor in the personality tending to dispose the individual toward a favorable or unfavorable adjustment to physical handicap. This study, of course, made no direct test of this contention, and such a test would be difficult to make. However, if contemporary circumstances affect the production of Barrier responses, it would seem that the duration and severity of illness would have influenced the Barrier score in the group studied. Since the results indicate that the Barrier response is probably independent of these variables, the overall observations at least seem congruent with the idea that the Barrier score represents a personality factor of a fairly basic nature.

The correlation found between adjustment and production of Barrier responses is quite consistent with results obtained by Fisher and Cleveland in a study of relationships between the Barrier score and certain personality variables (3). They found that the Barrier score in normals is related to such characteristics as greater participation in active sports, higher *n* Achievement scores, the production of TAT themes involving identification with a more definite image of the parents, and more emphasis on task completion. They concluded that emphasis on the Barrier aspect of body-image boundary is "a reflection of a style of life based on an unusually strong definition of self-identity and on objective self-expression aimed at establishing a stable, controlling relationship to the environment." This implies a degree of self-confidence that should serve as a bulwark against a threat to self-integration. The individual possessing these characteristics

obviously should be better able to absorb the threat to the self involved in physical loss or crippling and to integrate this loss into a revised self concept with the minimum of psychic trauma.

It should be noted that the body-image score has already proved meaningful in another crippled group with respect to the incorporation of loss. In an unpublished study by Fisher and Cleveland on data supplied by Haber (7), it was found that high-Barrier amputees dealt more realistically with phantom limb experiences than did low-Barrier amputees. One might conceptualize this as indicating that high-Barrier people were better able to integrate the body mutilation into their body-image schema.

These experimental observations lend support to the theoretical significance generally assumed for the body image as a factor in an individual's adaptation to physical handicap. It would seem, therefore, that the study of the individual's attitude toward his body is a promising way of approaching the problem of adjustment to disability. Meaningful relationships have been demonstrated between adjustment and the boundary aspect of body image. Perhaps it would be rewarding to extend investigations concerning physical handicap into other dimensions of the body concept.

SUMMARY

This study has shown, in a group of 56 poliomyelitis patients, a relationship between psychological adjustment to physical handicap and a measure of the boundary aspect of body image. This measure, derived from the Rorschach protocol, is known as the Barrier score. Apparently, individuals who conceive of their body boundaries as possessing defensive, armoring, barrier-like qualities are significantly better able to adapt to the circumstances surrounding physical loss than are those individuals whose body concept does not contain these features. The investigation of other dimensions of body image is suggested as a promising approach to the study of adjustment to handicap.

REFERENCES

1. BERGER, S. Paraplegia. In J. F. Garrett (Ed.), *Psychological aspects of physical disability*. Department of Health, Education, and Welfare, Office of Vocational Rehabilitation. Washington: U. S. Government Printing Office. Pp. 46–59.
2. FISHER, S., & CLEVELAND, S. E. The role of body image in psychosomatic symptom choice. *Psychol. Monogr.*, 1955, **69**, No. 17 (Whole No. 402).
3. FISHER, S., & CLEVELAND, S. E. Body-image boundaries and style of life. *J. abnorm. soc. Psychol.*, 1956, **52**, 373–379.
4. FISHER, S., & CLEVELAND, S. E. Relationship of body-image boundaries to memory for completed and incompleted tasks. *J. Psychol.*, 1956, **42**, 35–41.
5. GLUD, E., & BLANE, H. T. Body-image changes in patients with respiratory poliomyelitis. *Nerv. Child*, 1956, **11**, No. 2, 25–39.
6. GRAYSON, M. Concept of "acceptance" in physical rehabilitation. *J. A. M. A.*, 1951, **145**, 893–896.
7. HABER, W. B. Effects of loss of limb on sensory functions. *J. Psychol.*, 1955, **40**, 115–123.
8. LINDQUIST, E. F. *Statistical analysis in educational research*. New York: Houghton Mifflin, 1940.
9. MENNINGER, K. Psychiatric aspects of physical disability. In J. F. Garrett (Ed.), *Psychological aspects of physcial disability*. Department of Health, Education, and Welfare. Office of Vocational Rehabilitation. Washington: U. S. Government Printing Office. Pp. 8–17.
10. PRUGH, D. G., & TAGIURI, C. K. Emotional aspects of the respirator care of patients with poliomyelitis. *Psychosom. Med.*, 1954, **16**, 104–128.

Received September 19, 1956.

THE MANIFEST ANXIETY SCALE AS A MEASURE OF DRIVE

ROBERT E. SILVERMAN

New York University

THE Taylor Manifest Anxiety Scale (MAS) (16, 17) was designed to provide a technique for distinguishing drive level in humans in order to test aspects of Hullian learning theory. It was assumed that the scale measured differences in emotional responsiveness which contribute to the total effective drive state. According to Taylor (17), the title of the test should not be construed as suggesting special clinical validity since the scale was not designed for clinical purposes. In addition, Taylor and her associates have not been concerned with the relationship of anxiety per se to performance.

In spite of the original purpose of the MAS, many investigators have sought to relate the scale to clinical criteria of anxiety (3, 6, 8, 12); others have persisted in attempting to relate anxiety to performance variables (1, 2, 5, 11, 15, 18). Thus, many investigations owe their origins to Taylor's recent work and to Hullian theory. If these many studies have not furnished clear answers, they have served an heuristic purpose in highlighting the following questions: (*a*) What are the relationships between drive level and performance? (*b*) Does anxiety intensify effective drive level? (*c*) Does the Taylor MAS identify anxiety? (*d*) Does the Taylor MAS identify differences in drive level?

The first question continues to be extensively treated by students of Hullian theory, while the second and third questions await an unequivocal criterion of anxiety. The purpose of the present study was to consider the fourth question directly since the assumed relationship between the MAS and drive has been the focal point of a variety of recent investigations.

If one is willing to argue that drive may be defined solely in terms of the MAS, then the question becomes superfluous. However, it would seem that the intention of the Iowa group was to relate general drive variables to pe formance, not simply MAS scores to performance. In this light, the question becomes: What are the relationships between the MAS and measures of emotional responsiveness? If the MAS is a measure of emotionality, then, according to Taylor, it should be related either to a "chronic emotional state" in which subjects (*S*s) demonstrate high levels of emotionality or to differences in the potentiality for arousal, manifested by different degrees of emotional reactions to novel or threatening situations.

Palmar skin conductance and conductance changes were employed as measures of emotional responsiveness in the research to be reported. This technique, recently described by Schlosberg (14), has the multiple advantages of simplicity, reliability, and accuracy in estimating general states of autonomic nervous system activity.[1]

Because two recent investigations have failed to demonstrate a linear relationship between the MAS and the GSR or between the MAS and palmar perspiration technique (1, 9), steps were taken to explore the conditions that might obscure such relationships by employing two psychometric inventories. The two inventories were the MAS and a forced-choice version of the MAS designed by Heineman (7). The forced-choice inventory was included to provide some control over one of the known sources of error in the MAS, the influence of social desirability variables (7, 12).

METHOD

Subjects

The *S*s were 85 male undergraduate students from the elementary psychology course at the University College of Arts and Sciences at New York University. The *S*s were administered an inventory containing the Taylor MAS and the K scale from the Minnesota Multiphasic Personality Inventory as a part of the weekly small group discussions held in conjunction with the lecture course. The skin conductance readings were taken in a laboratory setting bearing no physical, temporal, or otherwise obvious relationship to the tests. In all, 66 *S*s participated in the MAS and the laboratory

[1] A study by the author, recently completed, indicates week-to-week reliability coefficients for the conductance measures ranging from .68 to .74. In addition, conductance changes were found to be systematically related to shock stress.

phases. Six weeks after the MAS was administered, the Heineman forced-choice version of the original Taylor scale (7) was given again in the discussion meetings, this time to 77 of the original *Ss*.[2] The attrition of *Ss* was due mainly to class absences and failure to keep appointments.

Apparatus and Procedure

The apparatus for measuring skin conductance consisted of a 50 microampere meter in series with two quarter-sized silver-silver, chloride-saline paste electrodes. This is essentially the unit outlined by Schlosberg (14). The circuit was run by a single one-and-one-half volt flashlight battery which, with the aid of a potential divider and a calibrated resistor, permitted absolute conductance to be read directly. In order to insure consistent electrode placement from *S* to *S*, and to maintain optimal contact during the experimental session, the electrodes were attached to *S* by having him grasp a cylindrical two-inch wooden bar on which the electrodes were seated in rubber cups. A second padded bar was then firmly clamped over each hand. Thus, *S* was in a kind of hand stock which permitted finger movements without disturbing the electrode contact.

Each *S* was introduced into the experimental setting by being placed in the stock and instructed to relax and rest while the experimenter (*E*) recorded his skin temperature. (This subterfuge was employed in order to obviate any concern on *S*'s part with electric currents and possible shock.) During this rest period of ten minutes, *E* recorded two conductance levels every 30 seconds, one from each electrode. This was done to insure the obtaining of the optimum conductance reading. Only the highest reading from each pair of readings was used in computing the mean conductance level for a given *S*.

Following this rest period, all *Ss* were instructed to perform a simple series of tasks consisting of 25 simple addition problems presented on a small screen directly in front of them. One group of *Ss* ($N = 36$), selected at random, was assigned to the shock-threat condition. They were fitted with an arm band containing two electrodes and were instructed that,

> During the next few minutes you are going to be shown some numbers on this screen. You are to add the numbers as quickly as possible . . . but also as accurately as you can. If you are not performing well, you will receive a shock . . . something like this.

(A sample shock from an Applegate Constant Current Stimulator set to deliver 2 milliamperes was given here.) The no-shock *Ss* ($N = 30$) were given the same task instructions but were not fitted with the electrodes and were not threatened with shock.

The *E* recorded two conductance readings immediately preceding the presentation of each problem. The entire problem series lasted 12½ minutes.

RESULTS AND DISCUSSION

The mean conductance scores for the rest periods were transformed into logarithmic units for the purpose of statistical comparisons.[3] The shock and no-shock task conditions provided measures of arousal which were computed in terms of the changes in log conductance from the rest periods to the task periods.

The correlations of the MAS scale and the forced-choice scale (FC) with the log conductance readings are shown in Table 1. None of the MAS correlations approaches significance. However, the FC scale does show a reliable if slight relationship with the rest period readings ($r = .24$) and a clear negative relationship with the conductance changes under the threat of shock ($r = -.50$). This negative relationship indicates that high scores on the FC scale are associated with low degrees of arousal.

An explanation for the fact that the two scales, the MAS and the FC, consist of similar items but do not correlate similarly with the conductance measures, may be found in Table 2 where the relationships among the MAS, the FC, and the K scales are shown. The K scale was designed to be a measure of "test-taking attitude" (10). High K scores are thought to be associated with a type of defensiveness against psychological weakness, while low K scores may indicate excessive self-criticism. Thus, the correlation of −.55 between the MAS and the K scale may be an indication that some low MAS scores were associated with defensiveness and some high MAS scores were simply reflecting a too candid or critical attitude. The FC scale, although correlating reliably with the MAS, did not show a relationship with the K scale.

These findings suggest that the failure to demonstrate a correlation between the MAS and the conductance measures may have been due to the spurious nature of some of the MAS scores. When steps were taken to minimize the effects of the test-taking attitude by employing a forced-choice technique, a relationship was found between the questionnaire and the physiological measure. Thus, an inventory like the Heineman scale appears to be a more sensitive measure of emotional responsiveness than the MAS.

It was observed above that conductance

[2] Heineman's FC-1 was used, scored according to Key 2 as described in the original report (7).

[3] In addition to being normally distributed, log conductance changes have been found to correlate well with measures of palmar perspiration. Mowrer *et al.* (13) reported a relationship between palmar perspiration, measured by a colorimetric technique, and measures of "psychological" tension during psychotherapy.

TABLE 1
CORRELATIONS BETWEEN THE TWO SCALES AND LOG CONDUCTANCE READINGS

	MAS		FC	
	r	*SE*	*r*	*SE*
Conductance: rest	.07	.11	.24*	.11
Conductance change: no shock	−.14	.18	.12	.20
Conductance change: shock	−.17	.16	−.50**	.14

* Significant at the .05 level.
** Significant at the .01 level.

TABLE 2
INTERCORRELATIONS AMONG THE THREE SCALES

	FC	K
MAS	.50** (*N* = 69)	−.55** (*N* = 85)
FC	—	−.07 (*N* = 69)

** Significant at the .01 level.

change is inversely related to the FC scale under the shock threat condition. This suggests that persons defined as anxious *by the scale* are less aroused by a threatening situation. These findings are in direct contradiction of the hypothesis that high levels of emotional responsiveness are associated with increased potentialities for arousal. Nevertheless, an interesting and not unfamiliar relationship is indicated which points up possible shortcomings in the conception of anxiety as a drive. Silverman and Blitz (15), using rote learning measures, found that high anxious *S*s (defined by MAS) were less affected by threat conditions than their low anxious controls. Deese, Lazarus, and Keenan (5) similarly reported that the rote learning of high anxious (MAS) *S*s was not reliably affected by shock stress. In both of these studies, the dependent variable, rote learning, did not offer any clue in regard to other behaviors possibly associated with threat and related to rote learning.

Yet it is unlikely that anxious persons are generally unresponsive to threat situations. It may be that the threat conditions of the laboratory are relatively benign by comparison to the personal doubts and fears which anxious persons habitually carry with them. Expressed in terms of this study, one might say that the tensions engendered by the laboratory threat did not raise an already high state of emotionality. A more general conception is that increases in arousal depend upon the already existing state of tension. This conjecture is supported by the fact that in the present study the amount of conductance change was a function of conductance level during rest. The product-moment correlation between rest level and change was .63. Darrow (4) also found that the size of the autonomic reaction (skin resistance) is a function of the amount of "available energy" which a given individual has at his disposal. Thus, it would seem that persons whose tension levels are high initially do not show much arousal and, hence, are unable to cope with novel situations which demand increases in arousal.

SUMMARY

This study was concerned with the relationship between the Taylor MAS and skin conductance. A group of *S*s was given the MAS and six weeks later the Heinman forced-choice version of the scale. Skin conductance measures were obtained for each *S* under two conditions: (*a*) a rest period, and (*b*) a task period involving shock threat for some *S*s, and no shock threat for others.

The results indicated that the forced-choice scale was positively related to the readings taken under rest condition, and negatively related to changes in conductance obtained under the threat of shock. The MAS did not correlate with any of the conductance readings.

These findings were explained by considering the high correlation between the K scale and the MAS and by noting that anxiety has been found by other investigators to be associated with low arousal.

REFERENCES

1. BEAM, J. C. Serial learning and conditioning under real-life stress. *J. abnorm. soc. Psychol.*, 1955, **51**, 543–551.
2. BITTERMAN, M. E., & HOLTZMAN, W. H. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615–623.
3. BUSS, A. H., WIENER, M., DURKEE, A., & BAER, M. The measurement of anxiety in clinical situations. *J. consult. Psychol.*, 1955, **19**, 125–129.
4. DARROW, C. W. The equation of the galvanic skin reflex curve: I. The dynamics of reaction in

relation to excitation background. *J. gen. Psychol.*, 1937, **16,** 285–309.

5. DEESE, J., LAZARUS, R. S., & KEENAN, J. Anxiety, anxiety-reduction, and stress in learning. *J. exp. Psychol.*, 1953, **46,** 55–60.
6. GLESER, GOLDINE, & ULETT, G. The Saslow-Screening test as a measure of anxiety-proneness. *J. clin. Psychol.*, 1952, **8,** 279–283.
7. HEINEMAN, C. E. A forced-choice form of the Taylor Anxiety Scale. *J. consult. Psychol.*, 1953, **17,** 447–454.
8. KENDALL, E. The validity of Taylor's Manifest Anxiety Scale. *J. consult. Psychol.*, 1954, **18,** 429–432.
9. LOTSOF, E. J., & DOWNING, W. L. Two measures of anxiety. *J. consult. Psychol.*, 1956, **20,** 170.
10. MCKINLEY, J. C., HATHAWAY, S. R., & MEEHL, P. E. The MMPI: VI. The *K* scale. *J. consult. Psychol.*, 1948, **12,** 20–31.
11. MALMO, R. S., & AMSEL, A. Anxiety-produced interference in serial rote learning with observations on rote learning after partial frontal lobectomy. *J. exp. Psychol.*, 1948, **38,** 89–101.
12. MATARAZZO, J. D. MMPI validity scores as a function of increasing levels of anxiety. *J. consult. Psychol.*, 1955, **19,** 213–217.
13. MOWRER, O. H., LIGHT, B. H., LURIA, X., & ZELENY, M. P. Tension changes during psychotherapy with special references to resistance. In O. H. Mowrer (Ed.), *Psychotherapy.* New York: Ronald Press, 1953. Pp. 546–640.
14. SCHLOSBERG, H. Three dimensions of emotion. *Psychol. Rev.*, 1954, **61,** 81–88.
15. SILVERMAN, R. E., & BLITZ, B. Learning and two kinds of anxiety. *J. abnorm. soc. Psychol.*, 1956, **52,** 301–303.
16. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48,** 285–290.
17. TAYLOR, JANET A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53,** 303–320.
18. WELCH, L., & KUBIS, J. The effect of anxiety on the conditioning rate and flexibility of the PGR *J. Psychol.*, 1947, **23,** 83–91.

Received September 19, 1956.

THE PSYCHOLOGICAL MEANING OF ACQUIESCENCE SET FOR AUTHORITARIANISM[1]

N. L. GAGE, GEORGE S. LEAVITT,[2] AND GEORGE C. STONE[3]

Bureau of Educational Research, University of Illinois

THIS paper is directed against the argument that the California F (predisposition to fascism) scale (1) and similar inventories have their validity lowered by their loading with acquiescence set. The argument runs as follows: (*a*) the F scale consists entirely of "negative" statements, i.e., statements that, when agreed with, reflect authoritarian attitudes; (*b*) the tendency to agree with ambiguous and unstructured statements has high internal consistency and generality over tests of markedly different content (7); (*c*) the F scale, and other inventories of attitudes about social issues and interpersonal relationships which employ agree-disagree items, are significantly loaded with irrelevant variance in acquiescence tendency.

Thus, Cohn (5) found the F scale significantly correlated (.41) with his Plus scale; the latter measures the number of times one responds "true" to 33 MMPI items dealing, according to Cohn, with nothing in particular. Similarly, Bass (4) found the F scale correlated positively with an acquiescence factor obtained by factor analysis.

A CORROBORATIVE STUDY

The present authors have obtained similar results in an independent study. We used Cohn's Plus Scale, the F scale, the Minnesota Teacher Attitude Inventory (MTAI) (6), and an especially constructed "Information-True" test of acquiescence. These were administered to 118 graduate students in education at the University of Illinois.

The Information-True test consisted of 50 "difficult" true-false information items. These items had been written and found through pretesting to be sufficiently difficult and obscure to elicit approximately a 50-50 split of "true" and "false" responses. The difficult items were mixed with 40 "easy" items; the latter were written and found through pretesting to be at about the 75 per cent level of difficulty. (The corrected split-half reliability of the number-correct scores on the easy items was .35; the latter score served as a crude measure of information about history, science, civics, Gallup Poll data, geography, and the like.)

The corrected split-half reliability of the acquiescence score, number of "true" responses on the difficult information items, was .68. This *r* indicated that the test was successful in eliciting reliable individual differences in tendency to respond "true," or acquiescence set. That difficult and ambiguous items are required to elicit the acquiescence set is again demonstrated by the fact that the reliability of the acquiescence score on the easy information items was only .09.

Table 1 shows the correlations among the Cohn Plus Scale, the Information-True score, the MTAI, and the F scale. These *r*s corroborate previous findings. The significant positive correlation (.19) between our Information-True score for acquiescence and Cohn's Plus Scale shows that acquiescence tendency is indeed general over considerably different types of test content. The correlation probably would have been higher if Cohn's Plus Scale were not concerned with another variable, namely, tendency to self-disparagement. That is, on most of Cohn's 33 MMPI items, the "true" response seems also to be the unfavorable description of oneself. This impression was supported by the judgments of two graduate students and two clerical workers who were asked to indicate whether "true" or "false" would constitute an unfavorable self-description in response to the 33 items. Their independent judgments agreed unanimously that "true" was the self-disparaging response to almost all the items.

Table 1 also shows that the Plus scale and the Information-True score correlate negatively (−.29 and −.26, respectively) with the MTAI. These negative *r*s were expected on the

[1] This investigation was supported in part by a research grant (M-650) from the Institute of Mental Health, National Institutes of Health, Public Health Service.

[2] Now at Fresno State College.

[3] Now at G. D. Searle & Co., Skokie, Illinois.

TABLE 1
CORRELATIONS AMONG ACQUIESCENCE, MTAI, AND F SCALE
(N = 118)

Measure	Mean	SD	r_{tt}	r's		
				(2)	(3)	(4)
1. Acquiescence: Plus Scale	12.7	5.1	.76†	.19*	−.29**	.44***
2. Acquiescence: Information-True	27.2	5.7	.68†		−.26**	.35***
3. MTAI	53.8	32.4	.86†			−.62***
4. F scale	49.8	23.8	.83‡			

* Significant at the .05 level.
** Significant at the .01 level.
*** Significant at the .001 level.
† Corrected split-half.
‡ Kuder-Richardson formula 20.

grounds that (*a*) the acquiescence tendency is general over highly diversified types of test content, and (*b*) the MTAI scoring weights are negatively correlated with degree of agreement to its items. The latter point is apparent from the published scoring keys for the MTAI. Of the 274 +1 scoring weights in the MTAI key, 60 per cent are assigned to "disagree" and "disagree strongly" responses, while only 21 per cent are assigned to "agree" and "agree strongly" responses. Hence, if acquiescence set is general over tests, acquiescent subjects should get low MTAI scores. Our two significant negative rs support this reasoning.

Table 1 further shows that both acquiescence tests correlate positively with the F scale. The r between the Plus scale and the F scale is .44; that between the Information-True score and the F scale is .35. These positive rs were expected by the same reasoning as that given for the negative rs between acquiescence and the MTAI. On the F scale, however, the scoring weights correlate *positively* (+1.00) with degree of agreement on the items. Our positive rs confirm those of Cohn and Bass.

The final r in Table 1 is that between the MTAI and the F scale. This r should be negative because both tests are affected by acquiescence tendency but score it in opposite directions, and because the content of the two tests is fairly similar. The obtained r was −.62.

PSYCHOLOGICAL SIGNIFICANCE OF ACQUIESCENCE

So far, our data merely support previous findings concerning the acquiescence set's internal consistency, generality, and effect on test intercorrelations. At this point, however, we can choose between two interpretations of these findings. That is, granted that the California F scale and the MTAI are loaded with acquiescence set, we can say either (*a*) that acquiescence is psychologically irrelevant to the authoritarian attitude measured by the F scale and the MTAI, or (*b*) that acquiescence is part and parcel of the authoritarian attitude. The latter position, developed by H. J. Leavitt and others (13), is that which we shall adopt, extend, and further test in the remainder of this study.

Not only the MTAI and the F scale are involved in this issue. As Bass (4) has indicated, measures of many other attitudes are implicated. These measures involve attitudes about social relations, morals, prejudice, custom, and status relations; examples are "dogmatism," "misanthropy," "xenophobia," and "suggestibility."

Positive and Negative Items

The act of agreeing. Let us assume that acquiescence tendency reflects a genuine aspect of the authoritarian disposition. Then agreeing with a statement should, other things being equal, reflect authoritarianism. But suppose the statement agreed with is one that expresses a nonauthoritarian attitude, i.e., is a "positive" statement. Examples of positive statements can be found in Bass's G scale, and in Ancona's $\frac{1}{F}$ Scale (2); both of these were made up of items opposite in meaning to the original statements of the F scale. Examples of these items are the following: "The artist and the professor are much more important to society than the businessman and the manufacturer." "Human nature being what it is, universal peace will come about eventually."

The MTAI contains 38 such "positive" items, where agreement may be considered to reflect a nonauthoritarian attitude. Examples from the MTAI are: "Minor disciplinary situations should sometimes be turned into jokes." "Unquestioning obedience in a child is not desirable." "There is too great an emphasis upon 'keeping order' in the classroom."

As shown in Table 1, the act of agreeing to positive items can occur either (*a*) "logically,"

TABLE 2
HYPOTHESIZED EQUIVOCALITY OF RESPONSE TO POSITIVE ITEMS AND UNIVOCALITY OF RESPONSE TO NEGATIVE ITEMS

Type of Item	Response	
	Agree	Disagree
Positive	Logical Meaning: Nonauthoritarian Psychological Meaning: Authoritarian	Logical Meaning: Authoritarian Psychological Meaning: Nonauthoritarian
Negative	Logical Meaning: Authoritarian Psychological Meaning: Authoritarian	Logical Meaning: Nonauthoritarian Psychological Meaning: Nonauthoritarian

because the person genuinely has the nonauthoritarian attitude which the content of the item represents, or (*b*) "psychologically," because he tends to acquiesce. Now, if acquiescence reflects obeisance to authority, i.e., a form of authoritarianism, then reasons *a* and *b* have opposite psychological significance. The "logical" and "psychological" reasons for agreeing are in a sense opposed to each other. Response to a nonauthoritarian statement may therefore be considered to have equivocal meaning for the personality of the subject involved.

Table 2 also shows that on negative items, however, agreeing reflects authoritarian attitudes unequivocally, i.e., both "logically" and "psychologically." Examples of negative items are all those of the F scale, such as: "People can be divided into two distinct classes, the weak and the strong." "Familiarity breeds contempt." "War and social troubles may someday be ended by earthquake or flood that will destroy the whole world."

There are 112 such negative items in the MTAI: "If the teacher laughs with the class in amusing classroom situations, the class tends to get out of control." "A child should be taught to obey an adult without question." "Discipline in the modern school is not as strict as it should be."

The act of disagreeing. On positive items, disagreeing may occur (*c*) "logically," because the person is authoritarian and so genuinely rejects the nonauthoritarian attitude which the content of the statement represents, or (*d*) "psychologically," because he is nonauthoritarian and so does not tend to acquiesce irrationally when confronted with a difficult or ambiguous statement. Here the person does not have the kind of obeisance to authority of the printed word which may be considered to reflect an aspect of the authoritarian personality. The nonauthoritarian is not negativistic; he simply does not agree as often on merely temperamental grounds, and his alternative therefore is to disagree more often. For positive items, again, reasons *c* and *d* have opposite significance.

On negative items, however, the act of disagreeing has the same significance for authoritarianism whether it is a genuine response to item content or a temperamental response representing autonomy, nonyielding, and the like. In short, agreeing or disagreeing with negative items has relatively univocal meaning for the authoritarianism of the person being tested.

Hypothesis. By this reasoning, we arrive at the hypothesis that *negative items have greater validity than positive items for the measurement of authoritarianism.* Psychologically, disagreeing requires more "ego strength," "emotional security," and "self-confidence," than does the act of agreeing. Coupled with the logical significance of a nonauthoritarian response, i.e., of disagreeing with a negative statement, the meaning of the response is clearly nonauthoritarianism. Psychologically, agreeing may mean yielding to the authority of the printed word or a similar submission to the power and plausibility of the social context in which the test is given. Coupled with the logical significance of a nonauthoritarian response, i.e., of agreeing with a positive statement, the resultant meaning of the response is relatively unclear; it may mean either authoritarianism or nonauthoritarianism. When both positive and negative items are similar in content and degree of ambiguity, the latter items should have higher validity for measurement of authoritarianism.

Method and Results

We have tested this hypothesis by correlating authoritarianism scores obtained separately with the positive and negative items of the MTAI, with ratings of teachers by their pupils. To test our hypothesis requires an external criterion of authoritarianism, unconfounded with the subject's acquiescence. As such a measure, we used pupils' mean ratings of their teachers on Leeds' "My Teacher" rating scale. This scale consists of 50 Yes—No—? items of the following type:

"Is this teacher often 'bossy'? Does this teacher force her ideas on the pupils? Is this teacher usually fair with the pupils? Does this teacher think she is always right and the pupils always wrong?" In previous studies (6), pupils' ratings of teachers on this scale have been found highly internally consistent over items and highly discriminating between teachers. Further, these mean ratings by pupils have been significantly correlated with experts' and principals' ratings of teachers as to their ability to maintain discipline, create a friendly classroom atmosphere, establish a feeling of security on the part of pupils, exert a stabilizing influence on the class, and develop the pupils' self-reliance. It seems defensible to use these ratings as a criterion of the authoritarian personality trends of the type at which the F scale and the MTAI are aimed. The authors of the MTAI cite *The Authoritarian Personality* as an influence on their description of the adjustment mechanisms responsible for undesirable teacher attitudes.

The subjects in this study were 97 of the 98 teachers of grades 4–6 in a midwestern city (10). The criterion ratings of the teachers were furnished by their 2800 pupils. The reliability of the mean rating of teachers was .91, as estimated with Horst's formula (11).

Table 3 shows the correlations of the scores on the positive and negative MTAI items with the pupils' ratings of the teachers. The r of the 38 positive items was .13. The 112 negative items were divided, by means of a table of random numbers, into three sets of 38, 38, and 36. This procedure made the scores on negative items comparable as to length of test with the scores based on positive items. These three sets of negative items correlated .30, .22, and .22 with the pupils' ratings. The mean r was .25.

The t value of the difference between rs of .13 and .25, using Hotelling's formula for the t of a difference between r_{13} and r_{12} (12, p. 278), is 1.64. Since the direction of the difference was predicted, a one-tail level of significance is justifiable; hence, the difference is almost significant at the .05 level. This test of significance did not take account of the fact that the value of r of .25 is a mean of three rs and hence is more reliable than a single r.

When all 112 negative MTAI items were combined, the score on them correlated .31 with the pupils' ratings. This r is higher than the r of .26 obtained with scores based on the entire 150-item MTAI. Thus, the score based only on negative items is more valid than the entire inventory.

TABLE 3

CORRELATIONS OF TWO TYPES OF MTAI SCORE WITH PUPILS' RATINGS OF TEACHERS (N = 97)

Type of MTAI Score	No. of Items	M	*SD*	r_{tt}†	r with Mean Rating of Teacher by Pupils
Positive Items	38	10.3	11.8	.75	.13
Negative Items					
Set A	38	14.8	10.2	.82	.30
Set B	38	17.7	9.3	.75	.22
Set C	36	14.2	8.9	.73	.22
All items	112	46.8	26.0	.91	.31
Total Test	150			.90	.26

† Corrected split-half.

DISCUSSION

The main point of this paper is that acquiescence may belong to the family of dispositions—including authoritarianism, obeisance to authority, and conformity—that have been identified in studies employing methods other than agree-disagree items. Accordingly, the agree-disagree format and high loading with negative items of the F scale and the MTAI make a positive contribution to the validity of these inventories.

On the basis of this reasoning, we hypothesized that, on positive items, the "logical" significance of the response might be contradicted by the "psychological" significance of the act of agreeing or disagreeing. Hence, such items should be less valid than negative items, in which the act of agreeing or disagreeing represents psychologically the same authoritarianism or nonauthoritarianism as the logical significance of the response denotes. Positive support for this reasoning was yielded by the correlations of the two types of items in the MTAI with pupils' ratings of their teacher on a variable akin to authoritarianism.

Our results resemble those described by Rundquist (15, pp. 142–143) and Cronbach (7, p. 406). Rundquist cited studies showing negative opinion items more valid in discriminating between men receiving and not receiving public relief, between employed and unemployed engineers, between problem and normal children. Cronbach found that number-right scores on False statements on True-False achievement tests were more valid than scores on True statements against criteria consisting of grades on other types of examinations.

These results can be assimilated to ours only if acquiescence is considered to reflect not only obeisance and authoritarianism but also maladjustment, low morale, and low intelligence. Granted such assumptions, we could prepare diagrams like Table 2 to show that responses to positive items are equivocal in terms of these variables, while responses to negative items are not. That these assumptions are plausible is indicated by some of Crutchfield's findings: "The assessment staff rating on 'intellectual

competence' correlates −.63 with conformity score, this being the highest relationship of any found . . . An 'ego strength' scale . . . correlates −.33 and a staff rating on 'leadership ability,' −.30 with conformity" (9, p. 194). Similarly, Ancona's (2) finding that acquiescence is inversely related to need achievement would support our assumption.

Further tests of this hypothesis should be made. They should employ both kinds of items—positive and negative. Scores for authoritarianism on both types of items should be validated against external criteria of authoritarian personality tendencies in which the acquiescence set is not involved. Similarly, measures of acquiescence should be correlated against measures of morale, ego strength, conformity, and intelligence.

If negative items prove to be more valid for the measurement of authoritarian personality tendencies, this will have practical implications for test construction. Either inventories should consist primarily of negative items, as does the present MTAI; or most positive items should be unscored; or scores on positive and negative items should be obtained separately and weighted empirically.

It appears, with the hindsight provided by the present data, that the authors of the F scale and the MTAI built better than they knew. No mention of the acquiescence variable or of its possible kinship with authoritarianism is made in the descriptions of either scale's construction. Levinson (1, p. 59) in describing the construction of the anti-Semitism scale, gave as his reasons for the use of negative items (*a*) they tend to be more discriminating, (*b*) they can be so phrased that they express subtle hostility . . . (*c*) they are more reasonably used for measuring receptivity to anti-Semitic ideology, (*d*) evidence showed variability of response rather than a set to agree or disagree consistently, and (*e*) the scale was valid against external criteria. None of this reasoning considers the psychological difference between agreeing and disagreeing when confronted with a difficult or ambiguous item. Until now, the argument *against* using only negative items has not been that they will produce a uniform set, but rather that they will reflect *individual differences* in such a set. Previous writers have shown that this does in fact occur. Our findings show that it not only occurs, but that it probably contributes positively and in a psychologically meaningful way to the validity of the F scale.

The present results also have theoretical significance for the nature of acquiescence, authoritarianism, and related variables. Our results provide some support for the conceptualization of acquiescence as akin to the "yielding" or "conformity" to group pressure studied by Asch (3) and Crutchfield (9). Crutchfield (9, p. 194) reported an r of .39 between conformity and the F scale. Rudin (14) found the F scale correlated about −.35 with freedom from contextual influence in perception of both social and nonsocial stimuli. If acquiescence does represent a kind of conformity—in this case to the "pressure" of the printed word or of the context in which a plausible statement is presented—then it does represent an aspect of authoritarianism.

In short, the present study may aid in understanding the psychological meaning of acquiescence set, apart from its logical and mathematical significance. Acquiescence has long been suspected of being a psychologically meaningful variable. It appears that, in measuring authoritarianism, we do have one of the potentialities to which Cronbach referred: "Some of the response-set variance is potentially useful, some of it is an interference with measurement. The problem for the tester is to capitalize on the effect of response sets where they are helpful to validity . . . ." (8, p. 19).

Further studies are needed to pin down the proposition that acquiescence set should be capitalized on, rather than shunned, in tests of authoritarianism, ethnocentrism, teacher attitudes, and the like.

SUMMARY

Previous workers have hypothesized and found significant correlations between the California F scale and acquiescence set (reliable individual differences in tendency to respond "agree" or "true" when in doubt). To corroborate these findings and extend them to the Minnesota Teacher Attitude Inventory (MTAI), we constructed an information test of acquiescence, consisting of 50 true-false items about obscure matters. This test, Cohn's Plus Scale of acquiescence, the F scale, and the

MTAI all correlated significantly with each other in directions expected on the basis of the acquiescence set.

The question was then raised, is acquiescence an aspect of authoritarianism such that variance due to acquiescence set is psychologically relevant to authoritarianism? By assuming that acquiescing reflects authoritarian personality tendencies, we can derive the hypothesis that "positive" items are less valid than "negative" items for measuring authoritarianism. Teachers' scores on these two types of MTAI items correlated differently, as hypothesized, with pupils' ratings of the teachers on variables similar to authoritarianism.

The hypothesis implies that the F scale, the MTAI, and similar inventories are advisedly constructed primarily of "negative" items. The psychological meaning of acquiescence resembles that of authoritarianism, conformity, low ego strength, and low intelligence.

REFERENCES

1. ADORNO, T. W. *et al. The authoritarian personality.* New York: Harper, 1950.
2. ANCONA, L. Indagine sulla natura psichica del "Response Set." La "motivazione al successo." *Arch. psicol. neurol Psichiat.*, 1954, **15,** 23–74.
3. ASCH, S. E. *Social psychology.* New York: Prentice-Hall, 1952.
4. BASS, B. H. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51,** 616–623.
5. COHN, T. S. The relation of the F scale to a response to answer positively. *Amer. Psychologist,* 1953, **8,** 335. (Abstract)
6. COOK, W. W., LEEDS, C. H., & CALLIS, R. *The Minnesota Teacher Attitude Inventory: Manual.* New York: Psychological Corp., 1951.
7. CRONBACH, L. J. Studies of acquiescence as a factor in the true-false test. *J. educ. Psychol.*, 1942, **33,** 401–415.
8. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt,* 1950, **10,** 3–31.
9. CRUTCHFIELD R. S. Conformity and character. *Amer. Psychologist,* 1955, **10,** 191–198.
10. DELLA PIANA, G. M., & GAGE, N. L. Pupils' values and the validity of the Minnesota Teacher Attitude Inventory. *J. educ. Psychol.*, 1955, **46,** 167–178.
11. HORST, P. A generalized expression for the reliability of measures. *Psychometrika,* 1949, **14,** 21–31.
12. HOTELLING, H. The selection of variates for use in prediction with some comments on the problem of nuisance parameters. *Ann. math. Statist.*, 1940, **11,** 271–283.
13. LEAVITT, H. J., HAX, H., & ROCHE, J. H. "Authoritarianism" and agreement with things authoritative. *J. Psychol.*, 1955, **40,** 215–221.
14. RUDIN, S. A. The influence of context on the perception of complex stimuli. Unpublished doctor's dissertation, Univer. of Illinois, 1955.
15. RUNDQUIST, E. A. Form of statement in personality measurement. *J. educ. Psychol.*, 1940, **31,** 135–147.

Received September 24, 1956.

CULTURAL INFLUENCE ON THE PERCEPTION OF MOVEMENT: THE TRAPEZOIDAL ILLUSION AMONG ZULUS[1]

GORDON W. ALLPORT AND THOMAS F. PETTIGREW

Harvard University

TRADITIONALLY, theories of the visual perception of movement—with which the present study deals—have been divided into two classes: 1. The *nativistic*, i.e., theories emphasizing the role of retinal and cortical functions relatively unaffected by learning, habit, experience, or meaning; and 2. The *empiricistic*, i.e., theories giving primary weight to the role of experience and learning.

For our purposes it is essential to subdivide empiricistic theories into two groups:

Cumulative habit: Stressing the effects of many types of early, remote, and generalized experience which by transfer or cross conditioning become a major determinant of the perception of movement. Toch and Ittelson (22, p. 199) state that "contemporary empiricism" favors this type of approach, offering its explanations of perceived movement in terms of "weighted averages of experiential sediments of all kinds acting inseparably."

Object connotation (meaning): Explaining perceived movement largely in terms of familiar objects. One sees continuous wing motion in an electric sign representing a bird in flight, although the stimulus actually occurs discontinuously in two or in three fixed positions. This theory would hold that our familiarity with birds in flight causes us to fill the gaps with perceived motion. A good statement of this theory of stroboscopic movement may be found in James (9). This author insisted that "perception is of definite and probable things." In explaining illusions, James leaned heavily upon their resemblance to familiar objects. In so doing he was merely rendering more concrete and specific Helmholtz's theory of "unconscious inferences" and Wundt's "assimilation" theory.

At the present time none of these theories can be defended in their exclusive purity. F. H. Allport (1) makes this fact abundantly clear. No empiricist, for example, can deny the native physiological substrate of the perception of movement nor its structural properties as represented by Korte's laws governing the phi-movement (13). Conversely, no nativist can deny the contribution that past experience may make to perceived movement. Wertheimer, for instance, admits that "it is certainly correct that past experience can influence the conditions (Verhältnisse) of stroboscopic vision" (23, p. 79). Von Schiller (21, p. 195) makes the suggestion—especially important for the present research—that attitudinal set and expectancy are particularly effective in determining the perception of movement in ambiguous (*alternativ*) situations.

Many authors who have worked with perceived movement, e.g., Neff (17) and Hall and Earle (7), have favored an eclectic view. They have concluded that native and experiential factors both contribute, as do likewise momentary set and the previous level of adaptation. And they allow, among experiential factors, for both *cumulative habit* and *object connotation* (meaning).

While the eclectic position is no doubt correct there is, as Toch and Ittelson insist (22), still a fundamental question concerning "the relative primacy or importance of learning processes and physiological conditions." Theorists do tend to lean in one direction or the other. The distinction between nativists and empiricists still exists. Among the former, one thinks of Wertheimer (23), Koffka (12), Metzger (14), Michotte (15); among the latter, of Helmholtz (8), Piéron (19), Ames (2), Wundt (24), Cantril (4), G. E. Müller (16),

[1] For their services as interpreters we wish to thank Kate Mdlala, Ethan Mayisela, Emmanuel Zuma, and Mr. September. For assistance in conducting the experiments we are indebted to Eric Barnett, Peter Hey and Ann Pettigrew. In making arrangements for the investigations we received kind cooperation from the Native Affairs Commissioner of the Province of Natal, from Dr. John Bennett of the Polela Health Center, Dr. Wolfgang Bodenstein of the Ceza Medical Mission, Max Marwick of the Institute for Social Research of the University of Natal, Dr. B. Gampel of the Natal Medical School, and Mr. J. Mngadi of the Lamontville Community Center; also from the Paramount Chieftainess, Thoko C. Zulu. For the opportunity to spend several months as members of the Institute for Social Research at the University of Natal, we are grateful to the Principal, Dr. E. G. Malherbe.

The original suggestion for the experiment grew out of conferences with Hadley Cantril for whose strategic advice we are grateful. We benefited likewise from discussion with F. P. Kilpatrick and E. Engel. The investigation received financial support from a grant from the Rockefeller Foundation to Princeton University for research in perception.

Kilpatrick (11), and Toch and Ittelson themselves (22).

THE CROSS-CULTURAL APPROACH

To gain light on this dispute psychologists have often asked, "How about primitive peoples?" If we can find a tribe or a culture where relevant past experience can be ruled out, we could then determine whether the perception resembles that of western peoples. If it does so, then the argument for nativism is presumably stronger. The first extensive attempt to apply this test was made by W. H. R. Rivers during the Torres Strait expedition in 1898 (20). Rivers presented to the island natives a whole array of visual illusions and compared their reports with western norms. For some of the illusions there were no appreciable differences; for others, the natives seemed on the whole less susceptible than westerners. While Rivers himself does not make the point clearly, his results seem to show that illusions involving object connotation (e.g., a European street scene) are far less compelling to the natives than are illusions having no such object connotation (e.g., the rotating spiral) (20, esp. pp. 130 f.).

It is not easy for western psychologists to visit primitive tribes, nor to conduct among them adequately controlled experiments. The present article, however, deals with one such attempt. But before we describe it, the theoretical point at issue should be made entirely clear: *We do not claim to be testing the merits of the nativist or empiricist positions directly.* For reasons that will appear in the course of our discussion, we do not believe that comparative perceptual studies on western and on primitive peoples can solve this particular riddle. *We claim only to have illuminated the part played by object connotation (meaning) in the perception of motion as over and against the part played by either nativistic determinants or cumulative habit.* Our experiment is *not* able to distinguish between the role of these last two factors.

The Rotating Trapezoidal Window

Before the days of Gestalt psychology it was customary to regard visual illusions as oddities, as exceptional experiences to be accounted for either in terms of nativistic or experiential constraints. Today, however, we make little distinction between illusions and veridical perceptions, since no illusion lacks veridical elements and no veridical perception is devoid of subjective shaping. So-called illusions are simply instances of perception where the discrepancy between impression and knowledge (whether the knowledge be the subject's or the experimenter's) is relatively striking. It is in such "looser" conditions of perception that theorists often seek to obtain light on the relative weight of factors entering into the normal perceptual process. The reasoning is not unlike that which leads psychologists to study exaggerated functions in psychopathology in order to obtain light on the same but less exaggerated functions of the normal mind.

Our experiment follows this logic, making use of the rotating trapezoidal window described by Ames (2)—a device that has been called "a dramatic masterpiece of ambiguous stimulation" (1, p. 276). The window (Figure 1) is so proportioned that as it rotates, the length of the longer edge is always longer on the retina than is the shorter edge (even when the shorter edge is nearer). The resulting perception is normally one of oscillation or sway; the observer apparently tending to keep the longer edge nearer to him. Instead of seeming to rotate, as it actually does, the window is seen to sway back and forth in an arc of 90 to 180 degrees.

An appended cube and rod add great interest to the illusion, since the perceived *rotating* of these objects conflicts sharply with the perceived *sway* of the window. In consequence, the cube is usually seen to detach itself and swing without support in a ghostly fashion in front of the window (for that period of time when the shorter edge, to which it is attached, is in fact nearer to the subject). Similarly, the rod bends, twists or "cuts through" the mullions in order to accommodate itself to the phenomenal oscillation. The observer finds the bizarre effect both amusing and inexplicable.

The window used in the present experiment is smaller than the original Ames window; length 13¼ inches, height of the long side 12½ inches, height of the short side 9 inches. Ames demonstrated that within limits the ratio of the sides of the trapezoid to one another cannot affect the illusion. The optimum speed of rotation Ames reports as 3 to 6 r.p.m. Our own motor driven window ran slightly less

than 2 r.p.m. The original Ames window had mullions dividing it into 15 frames; ours had 6 frames (probably more normal for a "window"). For certain comparisons, we employed also a true rectangular window, 12″ x 10½″.

The explanation Ames gives for the illusion maintains (*a*) that the observer, owing to familiarity with rectangular windows assumes *this* window to be rectangular; and (*b*) that owing to long experience with doors, windows, and similar objects, the observer has learned to interpret longer retinal stimulations as coming from nearer objects. Hence, the longer edge of the window is interpreted as being nearer, and the window is seen to oscillate rather than to rotate.

Ames (2, p. 14) gives a clearly empiricistic explanation with a leaning toward the object connotation version:

> In his past experience the observer, in carrying out his purposes, has on innumerable occasions had to take into account and act with respect to rectangular forms, e.g., going through doors, locating windows, etc. On almost all such occasions, except in the rare case when his line of sight was normal to the door or window, the image of the rectangular configuration formed on his retina was trapezoidal. He learned to interpret the particularly characteristic retinal images that exist when he looks at doors, windows, etc., as rectangular forms. Moreover, he learned to interpret the particular degree of trapezoidal distortion of his retinal images in terms of the positioning of the rectangular form to his particular viewing point. These interpretations do not occur at the conscious level, rather, they are unconscious and may be characterized as *assumptions* as to the probable significance of indications received from the environment.

It should be added that Ames does not insist that object connotation ("windowness") is the sole determinant of the illusion. He himself employed a variety of trapezoidal figures and discovered that even a plane surface of trapezoidal shape arouses the illusion of sway, though to a much less degree than does a "window frame" (2, p. 29).

The Hypothesis

In order to test the "object connotation" theory, we studied various groups of Zulu children (10–14 years old) in Natal whose own culture is virtually devoid not only of windows, but, to a surprising extent, of angles, straight lines, and other experiential cues that would presumably "cause" the illusion if it were wholly a product of experience. Our hypothesis therefore is:

Zulu children, provided they are unacculturated (amabinca) *will report the illusion of sway in the trapezoidal window* less *often than will urbanized acculturated Zulu children* (amabunguka) *or than white* ("*European*") *children.*

The Zulu Culture

Zulu culture is probably the most spherical or circular of all Bantu cultures, possibly the most spherical of all native African cultures (though it would be difficult to prove this contention). The word "zulu" means heavens or firmament, and the aesthetic ideal of round rather than angular styles affects native art, architecture, and speech.

Huts are invariably round (rondavels) or else beehive shaped, whereas in other Bantu tribes they are sometimes square or rectangular. Round huts arranged in a circular form with round stockades to fence in animals, constitutes a typical African homestead (kraal). Fields follow the irregular contours of the rolling land, and never seem to be laid out in the neat rectangular plots so characteristic of western culture. See Figure 2.

The typical Zulu hut has no windows, and no word for such an apperture exists. In the more primitive beehive grass huts, doors are merely round entrance holes; in the round mud huts doors are amorphous, seldom if ever neatly rectangular. Cooking pots are round or gourd shaped. In his studies among Zulus, L. Doob (6) finds that the less acculturated natives, relative to westernized natives, show a statistically significant preference for circles over squares when they are asked to choose between designs drawn in these shapes (personal communication to the authors).

It is commonly said in Natal that Zulus fresh from reserves cannot plow a straight furrow and are unable to lay out a rectangular flower bed. Such inability is of course overcome with experience and training, but the initial defect would seem clearly related to the circularity that is characteristic of life on the reserves and to the lack of familiarity with straight layouts.

Linguistically, the same bias towards circularity is seen. While it is possible to say "round" in Zulu, there is no word for "square." There is a word for "circle" but not for "rectangle." To speak of window, of square, or of

FIG. 1. THE EXPERIMENTAL SITUATION

FIG. 2. CIRCULARITY IN THE ZULU ENVIRONMENT

rectangle at all, a Zulu is forced to borrow these terms from Afrikaans or from English—provided he is able to do so.

The Subjects

The experiment required the use of two contrasting groups of subjects (*Ss*): those who had lived all or most of the their lives in western culture, and those who were unacculturated. Even in the Bantu reserves or in Zululand itself it is not possible to make certain that a resident does not know what a window is like. While schools, churches, and health centers are few and far between, they are nevertheless within the possible range of visitation by most native inhabitants, even children. Our experiments at Polela and Ceza took place in health centers, at Nongoma in a court house. The *Ss*, to be sure, were brought in from remote parts of the reserves by lorry, or came on foot; but they had at least this one-time acquaintance with a rectangular building and windows.

Still, it is possible to say that the experiment dealt with two widely contrasting groups in respect to the degree of experience they had had with western architecture and ways of life. Some members of the more primitive groups, for example, may never have seen windows with rectangular panes of glass prior to the actual experimental situation.

By using herd boys as *Ss*—mostly between 10 and 14 years of age (few of them knew their age exactly)—we were able to make certain that they had never been off the reserves and had never attended school. Boys of the same age comprised our urban control groups: one group of European boys at Greyville Community Center; another group of Bantu boys at the Lamontville Community Center in Durban. Most of these urban boys were attending school.

Our major experiment thus involved the following groups:

Group A	Urban European boys	(20 cases)
Group B	Urban African boys	(20 cases)
Group C	Polela Rural Africans	(20 cases)
Group D	Nongoma Rural Africans	(20 cases)

A rough indication of the cultural differences between the rural and urban groups lies in answers to the question asked at the end of the experiment about the rectangular window, "What does this look like?" The percentage saying "window" or "window frame" among the urban children was 88; among the rural, 45.

Procedure

The procedure involves four conditions, varying two factors bearing on the perception of the illusion: monocular vs. binocular viewing and distance from the stimulus object. Each *S* saw first the rectangular, and then the trapezoidal window in at least 3 full revolutions under each of the following conditions.

First trial: 10 ft.	binocular
Second trial: 10 ft.	monocular
Third trial: 20 ft.	binocular
Fourth trial: 20 ft.	monocular

It was thought that this order would impose the "hardest" condition first and therefore minimize the effects of suggestion. One might fear that if at 20 feet with one eye a subject easily perceived the illusion he might become accustomed to expecting oscillation in the trapezoidal window at closer distances and under binocular conditions. Conversely, of course, it might be argued that a subject who cannot perceive the illusion at ten feet binocularly would form an expectation that might prevent his obtaining it under easier conditions. We shall refer later to a control experiment (starting at 20 feet monocularly) designed to check on any suggestive effect that might arise from our order of presentation.

The experimenter (*E*) required the assistance of a second psychologist who kept records of the *Ss*' reports, also of an interpreter with all African *Ss*. Care was taken to prevent *Ss* who had finished the experiment from communicating with *Ss* who had not.

After being put at ease, the *S* gave his age (if he knew it) and his degree of education (if any). The *S* then sat in a chair placed at the proper distance from the window and was told to watch carefully the movement that he would see. After approximately three revolutions the *E* asked, "How does it seem to you to be moving?" Often the *S* spontaneously used his hands to indicate the motion until the *E* was satisfied whether a full rotation or a fluctuation was intended. The use of hand motion by the *S* proved to be fully convincing, for when he reversed the hand at precisely the right moment for the illusion to occur there could be no question concerning his experience. This device gave a useful check on the accuracy of the translator's report of the *S*'s verbal statements.

After obtaining a report for the rectangular window in each of the positions, the trapezoidal window was inserted in place of the rectangular, and the same method of report employed. In addition, the *S* was asked to tell whether the motion of the trapezoidal window was "like" that of the first window. This procedure served as a further check on the verbal description and hand report. In nearly all cases it was possible to record a clear and unequivocal judgment of the *S*'s perception. Less than three per cent of all judgments were listed by the *E* as "uncertain."

Whenever the illusion was reported for the first condition, the bar was inserted and the *S* asked, "How does the bar move?" and "Does the bar stay straight?" On occasional trials when the *S* had reported both the sway of the window and the bending of the bar, the cube was attached and the *S* asked to describe its motion. In these cases there was usually laughter (as with American *Ss*) and considerable confusion and difficulty manifested in describing so unreal and "spooky" a motion. Because of the difficulty of communicating concerning these complex phenomena we make no further systematic use of the cube and rod in the present study.

At the conclusion of the experiment, the *S* was asked what the rectangular window "looked like." He also stated his preference for one of two geometrical drawings presented to him in pairs (a circle, square, trapezoid). He then received a slight payment for his services (usually one pound of sugar or a candy bar and sixpence).

RESULTS

General results. Table 1 gives the results for the two unacculturated groups (Nongoma and

TABLE 1
NUMBER REPORTING ILLUSION
(Boys 10–14 Yrs. of Age, $N = 20$ in Each Group)

Condition	Nongoma Rural			Polela Rural			African Urban			European Urban		
	Yes	No	Uncertain	Yes	No	Uncertain	Yes	No	Uncertain	Yes	No	Uncertain
First condition (10′, both eyes)	3	17	0	4	14	2	13	7	0	11	9	0
Second condition (10′, one eye)	14	6	0	16	4	0	19	1	0	19	1	0
Third condition (20′, both eyes)	8	12	0	17	1	2	16	3	1	16	4	0
Fourth condition (20′, one eye)	18	2	0	17	2	1	18	2	0	19	0	1
Totals	43	37	0	54	21	5	66	13	1	65	14	1

Polela Reserves) and for the two districts within metropolitan Durban, African (Lamontville), and European (Greyville).

Combining all four conditions, there is a very significant tendency for the urban groups to report the illusion more often than the rural groups (corrected $2 \times 2\ X^2 = 15.34$; $p < .001$). This difference is most marked with the first condition (corrected $2 \times 2\ X^2 = 12.38$; $p < .001$). There is also a significant trend with the second condition (corrected $2 \times 2\ X^2 = 4.80$; $p < .05$) and a slight tendency with the third condition (corrected $2 \times 2\ X^2 = 1.87$; $p < .20$) for the rural children to observe the illusion less often than the urban children. Virtually no difference exists with the fourth, 20 feet and one eye condition.

Table 2 expresses the results in an alternative way. Since four conditions of presentation were used we can determine in how many of these four conditions on the average each of the cultural groups reported the illusion. For the two unacculturated groups combined, the illusion is reported in 2.425 of the four conditions, while for the acculturated groups the average is 3.275. This mean difference has high statistical significance ($t = 3.51$; $p < .001$).

It is evident from Tables 1 and 2 that city dwellers, whether Zulu or European, find the illusion somewhat more compelling than do rural ("primitive") natives. This tendency is especially pronounced at 10 feet with binocular vision—a condition when binocular cues of true depth (in this case, true rotation) are most plentiful. The reader will also note that the results for Polela (rural) stand somewhat between those for the city children and those from Nongoma (rural). At ten feet binocularly, they resemble those of Nongoma; at twenty feet binocularly, those of the city boys. Polela is, in fact, one hundred miles closer to Durban than is Nongoma which lies in the heart of Zululand. There is no doubt that the children in Polela have somewhat more familiarity with western architecture (specifically with windows) than do the children of Nongoma. The results (Table 2) correspond to a continuum of cultures: city children having a maximum of familiarity with western architecture, Nongoma children the least.

TABLE 2
DISTRIBUTION OF SCORES

Sample	Number of Yes's					Average
	4	3	2	1	0	
Nongoma	2	4	10	3	1	2.15
Polela	4	10	3	2	1	2.70
Urban African	12	4	2	2	0	3.30
Urban European	11	5	3	0	1	3.25
Total ($N = 80$)	29	23	18	7	3	2.85

Preference for circles. Following the experiment, all *S*s were shown drawings of a square, a trapezoid, and a circle (in pairs), and asked to express a preference. Table 3 indicates that those who expressed a preference for the circle (at least once in the two pairings) tend in the African groups to report the illusion *less* often. This tendency holds for all experimental conditions for all three African groups. The relationship is statistically significant, however, for only the binocular conditions. Circle-preferring Zulu children report the illusion significantly less often than the angle-preferring Zulus in Conditions 1 and 3 (corrected $2 \times 2\ X^2 = 3.89$; $p < .05$), but the difference in the monocular, second and fourth, conditions is not significant (corrected $2 \times 2\ X^2 = 0.18$, *n.s.*). There are no differences ap-

TABLE 3
PERCENTAGE OF CASES REPORTING ILLUSION AMONG SUBJECTS PREFERRING AND NOT PREFERRING CIRCLE

Condition	Combined African Groups $N = 60$		European Group $N = 20$	
	Preferring Circle $N = 39$	Not Preferring Circle $N = 21$	Preferring Circle $N = 12$	Not Preferring Circle $N = 8$
10′ binocular	28	43	58	50
10′ monocular	79	86	100	88
20′ binocular	59	86	83	75
20′ monocular	87	90	100	88
All conditions	63	76	85	75

proaching significance between the circle and noncircle-preferring European *S*s.

Let us assume that the aesthetic preference for circles may provide an index of the subjective closeness of the individual to Zulu culture (since it is, as we have seen, overwhelmingly a circular culture). If we do so we may say that this subjective closeness seems to predispose the *S* to resist the illusion. Stated in terms of transactional theory, rectangles and trapezoids have less functional significance for him. His preception of the window's rotation is accordingly more frequently veridical.

We have noted that this influence is significant only in the conditions involving *binocular* perception. A reasonable interpretation would be that cultural effects cannot easily change the basic demand character of the illusion monocularly perceived, but may do so when binocular conditions leave more latitude for choice and for interpretation among a greater number of cues.

This result then, so far as it goes, lends some weight to the contention that "cultural significance" is playing an appreciable part in determining the results.

Illusion with rectangle. Before viewing the trapezoidal window, every *S* in all four conditions first saw the rectangular window rotating. The purpose was to make sure that the sway (oscillation) reported for the trapezoid was judged to be *different* from the motion of the rectangular window. In most cases, indeed, the *S* was able to make the distinction clearly, indicating by gesture and by words that the rectangular window went "round and round" whereas the trapezoid oscillated.

There were cases, however, where the rectangular window was reported as oscillating. In fact, nearly one-third of the 80 *S*s reported such a phenomenon at one or more of the four conditions. The actual percentage reporting sway in the *rectangular* window at each of the four conditions is:

First condition	0
Second condition	8
Third condition	16
Fourth condition	28

It is conceivable that this curious and somewhat unwelcomed finding may be a result of a "suggestive" order of presentation. Thus, no *S* seeing the rectangle before the trapezoid under the first condition (10 feet binocularly) reported the phenomenon. And, with the exception of 3 cases, no *S* reported the rectangular illusion in the second, third, or fourth condition *unless* he had previously reported the trapezoidal illusion. Altogether, 81 per cent of our *S*s reported the illusion monocularly at 20 feet for the trapezoidal window, but only 28 per cent did so for the rectangular window under the same condition. In virtually all these cases the *S*s had grown accustomed to seeing oscillation at some previous stage with the trapezoid.

Pastore (18), however, finds that more than half of his 58 American college *S*s reported sway with the rectangle during a three-minute exposure, and at considerable distance from the window (where the retinal angle subtended by the two shapes may be subliminal). He does not tell whether the *S*s had grown accustomed to the sway of the trapezoid before they reported sway in the rectangle. We must leave this problem for the time being unsolved.

A CONTROL EXPERIMENT

In order to determine whether unwanted suggestive effects, caused by our order of presentation, were influencing the results at the optimal stage for the trapezoidal illusion (viz., 20′ monocularly) we simplified our procedure with entirely new *S*s. Urban and rural Africans served as before. To secure the latter, we visited the Ceza Medical Mission in Zululand, approximately 20 miles from Nongoma. Both Ceza and Nongoma are in the deepest part of the native reserves, over 200 miles north of Durban. Polela, as we have said, lies about 100 miles west of Durban and has more European influence (e.g., European-

style architecture). This fact, we repeat, seems to explain why, as Tables 1 and 2 show, the Polela *Ss* report the illusion somewhat more frequently than do *Ss* at Nongoma or Ceza.

In the control experiment, the *S* was asked to cover one eye. Sitting at twenty feet from the rotating trapezoidal window he then described its motion (both in words and by hand motion). Later he was seated at ten feet from the object and using both eyes described the motion, comparing it with the previous motion. Finally he was, as in the other groups, asked his preference for the circular, square, or trapezoidal figures. None of these herd boys had ever been to school.

For an urban control group we used a fresh population of Lamontville boys of the same age range.

Results

Table 4 gives the results.

These data are practically identical with those of the Nongoma and urban African samples cited previously. Again, the rural Zulu group reported the illusion significantly less often than the urban Zulu group in 10′ binocular condition (corrected 2×2 $\chi^2 = 11.53$; $p < .001$), but no differences appear at the 20′ monocular condition. This similarity of data proves that the order of presentation is not an important variable.

At the Ceza Mission Hospital we tested also a group of eleven expectant mothers, only one of whom had ever left the reserve. Eight reported the illusion at 20′ monocularly, two did not, and one was uncertain. None of the eleven, however, reported it at 10′ binocularly. These cases confirm the trend in all our tables that "primitives" are less able to perceive the sway in the trapezoidal window illusion under marginal conditions (i.e., at 10′ binocularly) than are city dwellers.

TABLE 4

NUMBER REPORTING ILLUSION IN CONTROL EXPERIMENT

Condition	Ceza (Rural) $N = 24$			Lamontville (Urban) $N = 21$		
	Yes	No	Uncertain	Yes	No	Uncertain
First condition (20′, monocular)	22	2	0	20	0	1
Second condition (10′, binocular)	2	18	4	14	7	0

TABLE 5

PERCENTAGE OF CASES REPORTING ILLUSION AMONG SUBJECTS PREFERRING TWO CIRCLES

Combined African Groups (Ceza and Lamontville); $N = 44$

Condition	Preferring Two Circles $N = 15$	Preferring Less than Two $N = 29$
20′ monocular	92	97
10′ binocular	27	38

Something should be said concerning the qualitative differences reported by *Ss* who first reported the illusion at 20′ monocularly and then again at 10′ binocularly. Not infrequently their reports at 10′ binocularly were "mixed," that is to say, sometimes they saw the oscillation and sometimes not. In every case, the *S* was asked to tell the "difference" if any existed between the movement seen at 20′ monocularly and that at 10′ binocularly. Most *Ss* claimed that there was a difference: sometimes the window at the closer distance seemed to move faster, sometimes in a bigger arc, sometimes even in the reverse direction. And often, as we have said, the reports at 10′ binocularly were "mixed—the subjects reporting sometimes a full rotation and sometimes a sway. We record "yes" to the illusion at 10′ binocularly if at any point in the experiment the *S* reports a clear oscillation. Since the same criteria were applied at both Ceza and Lamontville, no source of error is introduced.

If the reader is acquainted with the illusion he will no doubt recognize this ambiguity in the perception at 10′ when binocular cues are powerful evidence for true rotation, and yet the tendency to see illusory sway likewise exists. Because of this dual tendency we consider 10′ binocularly as a *marginal* condition for the illusion. What is important for our purposes is the finding that under such marginal conditions urban children, who are familiar with western architecture, report the illusion much more frequently than do herd boys on the Zulu reserves.

As in the major experiment our *Ss* expressed their preference, in three paired comparisons, for a circle, trapezoid, or square. Since only seven cases of the 45 failed to choose the circle at least once, we changed our criterion

TABLE 6
PERCENTAGE OF SUBJECTS RECOGNIZING STIMULUS AS A "WINDOW" WHO REPORT THE ILLUSION

Condition	Recognizing "Window" $N = 20$	Not Recognizing "Window" $N = 24$
20′ monocular	100	88
10′ binocular	55	29

slightly from that used in Table 3. We determined the occurrence of the illusion among those who chose the circle *twice* as compared with those who chose it only once or not at all. Table 5 shows that in this population likewise, *S*s who show a preference for circles tend—particularly in the binocular condition—to report the illusion somewhat less frequently than those who do not. The differences are not statistically significant but are in the same direction as those reported in Table 3. The implication of this finding, we repeat, seems to be that *S*s whose aesthetic preference lies with the circularity of their tribal culture are the more resistant to the illusion.

At the conclusion of the experiment the investigator showed each *S* the rectangular window and asked, "What is this?" (In order to make certain that the children would have an opportunity to say "window" if they perceived the resemblance, the question was asked in three different ways in the Zulu language.) To one or more of these three questionings 67 per cent of the urban children said "window," but only 26 per cent of the Ceza children gave the same reply. If we combine all cases who said "window" at both Ceza and Lamontville we find an appreciable, though not statistically significant, tendency for them to report the illusion more often than do children who did not recognize the windowness of the stimulus object (corrected 2×2 $\chi^2 = 2.81$; $p < .10$ when both conditions are combined). So far as it goes, this finding (Table 6) lends support to the object-connotation theory of the perception of movement, especially under the 10′ binocular condition.

DISCUSSION

Our most striking finding is that under optimal conditions (monocularly at 20 feet) virtually as many primitive Zulus report the trapezoidal illusion as do urban Zulus or Europeans. Taking this one partial result by itself we can say that the experiment supports either the nativistic or the cumulative habit theory. It does not by itself give us grounds for choosing between them.

Nativists might argue, for example, that whenever a longer and a shorter projection on the retina occur simultaneously the longer will assume a figure character and therewith a frontal position in the perception (other conditions being equal). Thus, some form of isomorphism obtains between retinal-cortical processes and the perception itself.

An empiricist with a "cumulative habit" preference might say that myriad ocular-motor adjustments from infancy have built up a dependable expectancy that longer projections on the retina will betoken nearer objects. One learns through repeated experience that longer retinal images of trees, cattle, people, stand for *nearer* objects (provided, of course, that one assumes such objects to be of equal size whether far or near from the eyes). It is not necessary for the *S* to have acquaintance with specific objects (in this case a window) in order to make a similar inference. The transfer effect is wide. Even the shadows painted on the rotating window are reminiscent of the *S*'s experience with shadows in nature. Old experiences automatically condition novel experiences even though the latter are only analogous.

One assumption that may play a decisive part in this case is the assumption of "right angularity." From earliest life the child is conditioned to the fact that perpendicular objects best withstand the force of gravity. Circular though his culture is, his basic frame of reference is still one of verticals and horizontals. Seeing an entirely new object (the trapezoidal window) he assumes unconsciously (no less than do people who are familiar with windows) that its shape is rectangular. Just like people in western culture he may make this assumption even if he "knows" that the object is not in reality rectangular. This assumption, together with the assumption that longer objects on the retina are usually nearer objects, would predispose him to perceive that the longer edge of the window is always nearer (thus inducing the perceived oscillation). No less than people in western culture he would fail to "correct" his assumptions of right-angularity and of long-edges-

being-near-edges by his "knowledge" of the trapezoidal shape of the stimulus.

Our major result is clearly not compatible with a narrowly conceived object-connotation theory. It is not necessary for the *S* consciously to assume that the object is a window in order to experience the illusion. True, as Table 6 suggests, the specific object connotation seems somewhat to favor the illusion, but it is clearly not the decisive determinant. Thus, for example, 88 per cent of those who did not consciously recognize the frame as a "window" nevertheless experienced the illusion at 20′ monocularly.

Yet, at the same time, our results show that object connotation cannot be disregarded. It also plays a part. Let us review the evidence:

1. Under all *suboptimal* conditions, as we see in Tables 1 and 4 (10 feet monocularly, and binocularly at 10 or 20 feet) there is a tendency for unacculturated *S*s to report the illusion less frequently than do the acculturated.

2. The *S*s who recognize the "windowness" of the stimulus object tend to report the illusion somewhat more frequently especially at 10 feet binocularly (Table 6).

3. African *S*s expressing a preference for circles (assumed here to indicate a subjective closeness to the rotund Zulu culture) tend to report the illusion less often than those expressing preference for angular figures (Tables 3 and 5).

We conclude that experience with, and identification with, western culture make it more likely that the illusion will be perceived under marginal (suboptimal) conditions.

One fact reported by Ames, and mentioned above, supports our interpretation. He finds that a plane trapezoidal frame yields appreciable oscillation, but that the addition of mullions, panes, and shadows enhances the illusion. In other words, specific "thingness" contributes to the experience though it does not account for it wholly (6, pp. 28–31). And we may again allude to von Schiller's contention that expectancy is effective in determining perceived movement under marginal (*alternativ*) conditions.

May Brenner (3) likewise makes the point that when marginal conditions obtain, the *S* is forced to depend on stimulus *meaning*. On the other hand, when optimal stimulus conditions obtain, even brain-damaged cases report apparent movement to much the same degree as do normal cases.

Several other experiments have dealt with the effects of meaning on apparent movement. Thus, Jones and Bruner (10) report that in a stroboscopic experiment the line drawing of a man is seen to be in motion more actively than is a nonsense figure. De Silva (5) had previously established this same fact. Jones and Bruner conclude that "the more probable and practiced the movement, the more adequately will the movement experience maintain itself under suboptimal conditions" (10, p. 165). This conclusion is in agreement with our own.

Toch and Ittelson (22) report an experiment in which drawings of a bomb stroboscopically presented are seen in a downward (falling) motion, whereas drawings of an airplane presented in an identical fashion are seen in an upward (rising) motion. Though this experiment taken by itself favors an object-connotation theory, the authors argue in general for the cumulative-habit theory. They contend, rightly no doubt, that the nativist position cannot be adequately tested short of a longitudinal study of infants from birth. They believe that generalized past experience accounts for our major dispositions to perceive stroboscopic or other illusory movement, but allow that under conditions of ambiguity or equivocation specific meaning connotations will enter to determine the direction and nature of the movement. Here, too, our findings are concordant.

If we leave the field of experimental testing for a moment, we can find many familiar instances of the role of object connotation in resolving perceptual ambiguities. A streak of light in the night sky may be seen as a shooting star, as distant fireworks, or as a jet plane, depending largely on one's expectations. Bartlett tells of the Swazi chieftain who perceived all traffic policemen in London as friendly beings, because in Swazi culture the upraised arm is an amiable greeting. A child in a dentist's chair, more familiar with space-ships than with nitrous oxide, perceives the inhalator as a spaceship toy. Every projective test assumes that ambiguous (multivalent) stimuli will receive subjective structuring on the basis of need, set, expectancy, or habit.

Our experiment does not introduce factors of need or of set, but deals only with the relevance of past experience (meaning) as a determinant of perceived movement. It may, however, be pointed out that among sophisticated observers of the trapezoidal illusion under marginal conditions (e.g., at 10 feet binocularly) a volun-

tary effort to see or not to see the window as oscillating (or as rotating) can also be effective, especially if the observer picks out some detail of the window to watch during the rotation, thus inhibiting the impression as a whole. Meaning is not the only determinant entering into the resolution of perceptual ambiguity, but it is one of them.

Returning to James's statement that, "Perception is of definite and probable things," we may say that under optimal conditions of stimulation definite structure is conferred by physiological conditions or by deeply ingrained functional habits of spatial adjustment, or by both. But when marginal conditions prevail, an association with the most "probable" object is often called upon to provide the definiteness that is otherwise lacking.

What we have called "marginal" conditions should receive a further word of explanation. We use the term in our experiments to indicate that perceptual conflict is present. Under binocular conditions (especially at 10′) there are many cues that "give away" the true rotation; at the same time there are operating also the assumptions that the window is rectangular and that longer objects are nearer. Under such a condition of conflict our finding is that urban children resolve the conflict with the aid of the supplementary assumption of "windowness." Not being able to draw on this supplementary assumption, the rural children as a rule resolve the conflict in favor of the binocular (or true) evidence. In this particular case, therefore, one might say that the primitive children see things "as they are" more often than do the children of civilization.

CONCLUSION

The perception of motion as represented in the rotating trapezoidal window is governed, under *optimal* conditions, by nativistic determinants or by the unconscious utilization of residual (but not immediately relevant) experience, or by both. (Our experiment does not enable us to decide this issue.) At the same time, object connotation (meaning) based on closely relevant cultural experience helps to determine the nature of the perceived movement under *marginal* conditions.

An adequate theory of perceived movement must therefore allow a place for the subject's specific assumptions of meaning even though it cannot be based solely on this foundation.

REFERENCES

1. ALLPORT, F. H. *Theories of perception and the concept of structure.* New York: John Wiley, 1955.
2. AMES, A., JR. Visual perception and the rotating trapezoidal window. *Psychol. Mongr.*, 1951, **65,** No. 7 (Whole No. 324).
3. BRENNER, MAY W. The effects of brain damage on the perception of apparent movement. *J. Pers.*, 1956, **25,** 202–212.
4. CANTRIL, H. *The "why" of man's experience.* New York: Macmillan, 1950.
5. DESILVA, H. R. An experimental investigation of the determinants of apparent visual movement. *Amer. J. Psychol.*, 1926, **37,** 469–501.
6. DOOB, L. An introduction to the psychology of acculturation. *J. soc. Psychol.*, 1957, in press.
7. HALL, K. R. L., & EARLE, A. E. A further study of the pendulum phenomenon. *Quart. J. exp. Psychol.*, 1954, **6,** 112–124.
8. HELMHOLTZ, H. VON. *Physiological optics.* Vol. III, Chaps. 26, 78. Transl. by J. P. C. Southall. Optical Society of America, 1925.
9. JAMES, W. *Principles of psychology.* Vol. II, Chap. 19. New York: Henry Holt, 1890.
10. JONES, E. E., & BRUNER, J. S. Expectancy in apparent visual movement. *Brit. J. Psychol.*, 1954, **45,** 157–65.
11. KILPATRICK, F. P. Two processes in perceptual learning. *J. exp. Psychol.*, 1954, **47,** 362–370.
12. KOFFKA, K. *Principles of Gestalt psychology.* New York: Harcourt, Brace, 1935.
13. KORTE, A. Kinematoskopische Untersuchungen. *Z. f. Psychol.*, 1915, **72,** 193–206.
14. METZGER, W. *Gesetze des Sehens.* Frankfurt: Kramer, 1936.
15. MICHOTTE, A. *La perception de la causalité.* Paris: Vrin, 1946.
16. MÜLLER, G. E. *Komplextheorie und Gestalttheorie.* Göttingen: Vandenhoeck & Ruprecht, 1923.
17. NEFF, W. S. A critical investigation of the visual apprehension of movement. *Amer. J. Psychol.*, 1936, **48,** 1–42.
18. PASTORE, N. Some remarks on the Ames oscillatory effect. *Psychol. Rev.*, 1952, **59,** 319–323.
19. PIÉRON, H. Remarques sur la perception du mouvement apparent. *Année Psychol.*, 1934, **34,** 245–48.
20. RIVERS, W. H. R. *Cambridge anthropological expedition to Torres Straits.* Reports, Vol. II, Pt. 1: Introduction and Vision. Cambridge Univer. Press, 1901.
21. SCHILLER, P. VON. Stroboskopische Alternativversuch. *Psychol. Forsch.*, 1933, **17,** 179–214.
22. TOCH, H. H., & ITTELSON, W. H. The role of past experience in apparent movement: a revaluation. *Brit. J. Psychol.*, 1956, **47,** 3, 195–207.
23. WERTHEIMER, M. *Experimentelle Studien uber das Sehen von Bewegung. Drei abhandlungen zur Gestalttheorie,* Erlangen: Verlag d. Phil. Akad., 1925, 1–105.
24. WUNDT, W. *Gründzuge der physiologischen Psychologie,* Vol. II, (6th ed.) Leipzig: Engelmann, 1910.

Received April 23, 1957. Prior publication.

A MOTIVATIONAL FACTOR IN CONFORMITY[1]

GEORGE MOELLER AND MORTIMER H. APPLEZWEIG

Connecticut College

CONFORMITY in group situations may arise from many sources. This is a report of research into the effects on conformity behavior of individual motivation for social approval.

As we have indicated elsewhere (2, 3, 4), any of a number of behavior modes may serve as means toward satisfaction of a particular motive, and, conversely, a particular mode may serve any of a number of motives. This point of view implies that prediction of an *individual's* mode of behavior in a given situation must be based on knowledge of the situation and of the individual's motivation. Conforming behavior, as an example of one such mode, might then be predicted from a knowledge of individual motivation patterns, provided the performance situation permits gratification of the relevant motives in this way.

The present study was designed as a test of this general hypothesis. It was predicted that college students with high motivation for social approval would be more likely to agree with the judgments of their peers in the type of conformity situation developed by Asch (5, 6) than would students who were not so motivated. This hypothesis in no way precludes the possibility that conformity behavior may result from other patterns of motivation in other situations, but, rather, states a specific relationship between the strength of one motive and the occurrence of one mode of behavior in a particular situation.

METHOD

The Measure of Motivation

A new forced-choice motivation questionnaire, *The Behavior Interpretation Inventory* (9), was administered as part of the regular freshman orientation testing program at Connecticut College in the Fall of 1954. The inventory consists of 59 items, the stem of each representing a small sample of behavior, with four "reasons" given as alternatives. The four alternatives for each item are specific to the item, but are always some form of (*a*) escape from present unpleasantness, (*b*) concern for a secure future, (*c*) attainment of social approval, and (*d*) consistency with a self-picture. These four general motive categories were selected after a logical analysis of systems of classification of motives by Maslow (8), Murray (10), and Stein (15). A distinction was drawn between specific behaviors (modes) and generalized goal orientations (motives). Behaviors often grouped under the category of achievement motivation, for example, were considered to represent more appropriately behavior modes than underlying motives. A particular mode (e.g., acquisition, aggression, affiliation, etc.) can serve any one or more of an individual's motives, depending upon the situation in which it is called forth.

The utility of these particular motive categories was checked empirically in a study using an open-ended (sentence completion) form of *The Behavior Interpretation Inventory* (1). The four categories were found sufficiently inclusive to permit scoring of all response themas in the protocols obtained. Independent evidence for the usefulness of this motive classification system can be inferred from a recent study of men under military combat stress (11) where a list closely approximating our categories was obtained. The four choices likewise resemble the four "wishes" of Thomas (17).

Subjects (*Ss*) were asked to rank each set of four alternatives in the order of likelihood that they were the reasons for the behavior described in the item stem. Since the item stems represent a variety of situations and behavior modalities, the frequency of preference for a given motive was taken as an index of the strength of the motive relative to that of the other motives for a particular individual. Score on a given scale, then, was the total number of first choices assigned to alternatives appropriate to that scale.[2] A total of 25 behavior modes was used in the 59 items of the inventory, covering a wide range of situations and types of action. The two examples below represent modes of counteraction (*#* 35) and physical sentience (*#* 52). In brackets following each alternative is the number representing the motive category (see above).

35. He has been in poor health for some time, but, despite this, he is engaging in strenuous physical activity because:

(A) He fears his health might get even worse if he should take it easier [2].

(B) This relieves some of his tensions and lets him sleep better [1].

[1] This article constitutes Technical Report No. 5, Project NR 172-228, Contract Nonr 996(02) between the Group Psychology Branch, Office of Naval Research and Connecticut College. Permission is granted for reproduction in whole or in part for any purpose of the United States Government. A report of this research was given at the meetings of the Eastern Psychological Association, New York City, April, 1957.

[2] Alternative scoring for the sum of ranks has been used and is described elsewhere, as is the Inventory and its rationale in greater detail (2, 3, 4, 10, 16). The inventory was originally developed by the authors jointly with Drs. Dee G. Applezweig and Ross R. Thomas. Dr. Harvey Burdick participated in the revisions which resulted in the present form.

TABLE 1
MEAN BII z-SCORE PROFILES FOR SOCIAL-APPROVAL, SELF-APPROVAL AND SOCIAL- AND SELF-APPROVAL GROUPS

Group High	N	BII Scale: 1 Escape	2 Security	3 Soc. Appr.	4 Self Appr.
on					
Social Approval	15	−0.51	0.07	1.27	−0.51
Self Approval	13	−0.51	−0.73	−0.46	1.10
Social & Self Approval	13	−0.78	−0.56	0.37	0.67

(C) He has always believed in keeping one's body strong [4].

(D) He hopes that others will admire him for trying [3].

52. This man is lolling in the warmth of the water at a hot-springs resort. He often comes here because:

(A) He finds his mind is clear and he can think best when he is relaxed [4].

(B) He feels it is socially fashionable to vacation here [3].

(C) The sulphur of the water and the warmth relieve his rheumatism [1].

(D) He hopes that this will help him stay healthy longer [2].

Subjects

BII scores of 263 *Ss* in the class were divided into four groups on the basis of scores on the scales of social and self approval (scales 3 and 4, respectively). Individuals were classified as high and low on each scale as their scores fell above or below the median on that scale. Scores on the two irrelevant scales (1 and 2) were used as a further restriction on sampling to produce groups with approximately equal means on these two scales. A test sample of 88 *Ss* was finally obtained, consisting of *Ss* high on social approval (scale 3) and low on self approval (scale 4), *Ss* high on self approval and low on social approval, and *Ss* high on both scales. Of the 88 *Ss* run in the experiment, 41 were used, the remainder being discarded for the following reasons: 3 for error in the conduct of the experiment, 6 for error on the part of a member of the majority, 10 who guessed the true purpose of the experiment while participating in it, and 28 who appeared on subsequent interview to have been overly suspicious of the procedure and/or purpose of the experiment.[3]

Table 1 presents mean BII z-score profiles for the three groups whose data compose this study. It should be noted that the group high on both scales is actually not as high as either of the other groups on the relevant scale. This matter will be discussed further below.

Procedure

The experimental procedure followed that employed by Asch in his studies of the effect of social pressure on the judgment of length of lines (5, 6), with the exception that the size of the majority used for the present study ranged from 4 to 6, in most cases being 5, while Asch used 6 to 8 member groups.[4]

Each *S* was asked to participate in a study of perception. On reporting, *S* found 4 to 6 of her classmates (who had ostensibly received identical requests) waiting for the experimenter (*E*). The group was brought to the experimental room and seated in a semicircle facing the blackboard. *E* then read the following instructions:

> This is a task involving the discrimination of lengths of lines. Before you is a pair of cards. On the left is a card with one line; the card on the right has three lines differing in length; they are lettered A, B, and C, in order. One of the three lines at the right is equal to the standard line at the left—you will decide in each case which is the equal line. You will state your judgment in terms of the letter of the line. There will be 18 such comparisons in all.
>
> As the number of comparisons is few and the group small, I will call upon each of you in turn to announce your judgments, which I shall record here on a prepared form. Please be as accurate as possible. Do not talk to each other during the experiment. Any questions? Suppose we start at the right (*Ss'* right) and proceed to the left around the table. Will you begin, please?

The experiment was then conducted in accordance with the instructions. The 4 to 6 classmates of *S* had been instructed beforehand to give specific erroneous responses on 12 of the 18 trials. To reduce the possibility of error by the majority, the first member was given a list of the correct responses which she consulted inconspicuously when necessary; all other members of the majority repeated the leader's response. The majority was instructed to respond as though their judgments were truly their own, but not to overact their parts. In all cases the members of the majority took their seats so that all but one member was to *S*'s right, the other to her left.

At the conclusion of the experiment, *E* asked the members of the group, in turn, to report to separate rooms for a brief interview. *S* was, of course, the only member of the group actually interviewed. This interview served to inform *S* of the true nature of the experiment and thus relieve some of the anxiety which may have been generated in the test situation. It also provided information regarding the set *S* had taken, particularly whether or not she had accepted the instructions and the experiment at face value. The authors alternated as interviewer as their schedules permitted.

[3] Interviews were tape-recorded and reviewed by the authors without knowledge of the motivation-group to which the *S* belonged. Conformity score was likewise ignored in determining whether the data for a particular *S* would be discarded or not. Actually, the conformity scores of the 38 *Ss* discarded (for knowledge or suspicion of the purpose of the experiment) varied considerably, although the mean number of yields for this group was smaller than for those *Ss* used.

[4] Mrs. Joan Lake Kaiser served as experimenter in the conformity measurement procedures. We wish to express our thanks to her and to the many students who served as experimental assistants and subjects.

RESULTS

Quantitative

Figure 1 shows the percentage of *Ss* in each group which yielded to the majority on the twelve trials in which the majority responded incorrectly. It can be seen that the social-approval group consistently yielded more frequently than did either of the other two groups. The performance curves of the self-approval group and the group high on both scales could not be distinguished from each other. Further, from the third critical trial forward, the performance of the three groups seemed to reach a stable level, such differences as appeared neither increasing nor decreasing consistently.

Table 2 shows the number of *Ss* who yielded to, or compromised with, the majority a given number of times for each of the three groups. It can be seen that the frequency distribution for the social-approval group is somewhat skewed toward high frequency of yielding; while the distributions for the other two groups are skewed in the opposite direction. The Kruskal-Wallis nonparametric measure (H) for comparing sum of ranks (14) yielded a p close to .05 for these data, indicating that the groups do in fact differ in the predicted direction. Comparison of the several pairs of treatments by U tests further showed that the social-approval group differed from the other two, p being less than .05 in both cases (single-tailed test). The other two groups do not differ from each other.

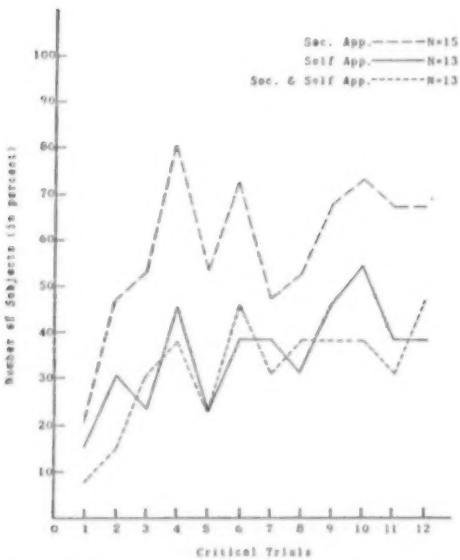

FIG. 1. PERCENTAGE OF CONFORMING AND COMPROMISE RESPONSES FOR EACH PROFILE GROUP ON THE TWELVE CRITICAL TRIALS

TABLE 2

FREQUENCY OF COMPROMISE OR YIELDING IN EACH OF THE EXPERIMENTAL GROUPS ON THE TWELVE CRITICAL TRIALS

Frequency	Experimental Groups		
	Soc. Appr. (Scale 3)	Soc. & Self Appr. (Scales 3 & 4)	Self Appr. (Scale 4)
	$N = 15$	$N = 13$	$N = 13$
12	1	—	1
11	1	2	—
10	1	1	—
9	2	—	—
8	3	—	2
7	1	—	1
6	2	1	1
5	1	—	2
4	—	2	—
3	1	1	—
2	1	—	2
1	1	1	—
0	—	5	4

Qualitative

The postexperimental interview yielded qualitative data that generally support the findings reported for the experiment proper. The standard questions that elicited different responses from the several groups, or that are of interest in themselves, are shown in Table 3, along with the frequency of the various responses. Data for the two *Ss* who always agreed with the majority are not included since most of the questions could not be asked of them.

Only Questions 1 and 7c evoked answers requiring any kind of clinical evaluation in scoring. These answers were typed verbatim, then analyzed without reference to the *Ss*' group membership. The criteria for classification are presented below along with appraisal of the answers. In no case did the analysis go beyond the explicit content of *Ss*' statements.

Responses to the first question were classified by the authors as intropunitive or nonintropunitive, as this concept is defined by Rosenzweig (12). Intropunitive responses were those in which *S* expressed doubt about the accuracy

TABLE 3

RESPONSES IN THE POST-EXPERIMENTAL INTERVIEW

Question 1:

Would you describe in your own words your experiences during this experiment?

Response Classified as:	Group		
	Soc. App.	Soc. & Self	Self App.
Intropunitive	8	3	5
Nonintropunitive	6	10	7

Question 2a:

When you gave an estimate that disagreed with the others, did you feel that if the lines were measured with a ruler you would turn out to be right or wrong?

Response	Group		
	Soc. App.	Soc. & Self	Self App.
Right	10	12	12
Wrong or ?	4	1	0

Question 2b:

Then did you feel that if the lines were measured, the others would turn out to be wrong?

Response	Group		
	Soc. App.	Soc. & Self	Self App.
Wrong	7	12	10
? or Right	7	1	2

Question 3:

Would you say that you were concerned about the disagreements?

Response	Group		
	Soc. App.	Soc. & Self	Self App.
No	5	8	7
Yes or ?	9	5	5

Question 4a:

Still thinking back to the time that you were giving your judgments, would you say that the others made you doubtful about your accuracy?

Response	Group		
	Soc. App.	Soc. & Self	Self App.
Yes	11	7	5
No	3	6	7

TABLE 3—*Concluded*

Question 5a:

Would you say that you were tempted at times to answer as the others did?

Response	Group		
	Soc. App.	Soc. & Self	Self App.
Yes	12	7	5
No	2	6	7

Question 6a:

Did you ever answer as the others did against your own first choice?

Response	Group		
	Soc. App.	Soc. & Self	Self App.
Yes	7	4	2
No	7	9	10

Question 7b:

Did you wonder what everybody might be thinking of you when you disagreed?

Response	Group		
	Soc. App.	Soc. & Self	Self App.
Yes	10	9	9
No	4	4	3

Question 7c:

What did you think the others thought about you?

Response Classified as:	Group		
	Soc. App.	Soc. & Self	Self App.
Something wrong with me	7	6	1
Other	3	3	8

of her judgment, e.g., "I think when she took them away I kept wondering, did I answer correctly . . . they were so close sometimes." A response was classified as nonintropunitive if *S* stated that the group was in error, or that some physical aspect of the situation led to the disagreements, or she simply reported that there were disagreements. An example of the atter type of response is, "Everybody was saying the line was A, and I said it was B. No-one else ever varied from their observations, but I did." The content of these examples is

typical of those assigned to the two categories, but they are unusually brief.

The responses to Question 1 of the social and self-approval group were more often non-intropunitive than were those of the other two groups. In response to Questions 2a to 6a, the social-approval group differed in the expected way from the other two groups. Members of that group were: less certain that they were right (Q 2a), less willing to state that the majority was wrong (Q 2b), more concerned about disagreements with the majority (Q 3), less certain about the accuracy of their judgments (Q 4a), more often tempted to answer as the others did (Q 5a), and more willing to admit that they had answered against their own first choice (Q 6a). None of these differences is completely convincing when taken alone, but in the aggregate they are consistent with the hypothesis that members of the social-approval group were more concerned over their performance relative to that of the majority than were members of the other two groups.

The differences that appeared in the interviews do not simply reflect differences in performance in the experiment proper. If we compare the performance scores of persons who said "Yes" and "No" (or "Right" and "Wrong") in reply to a particular question we find no differences. This lack of relationship between performance and *S*'s reaction to that performance arises, of course, because *S* may yield uncritically, yield but feel concern, not yield and feel concern, or remain independent in both performance and attitude.

The responses to Questions 7b and 7c clearly reflect this distinction between behavior and the concomitant affect. Data for Q 7b are presented to show that awareness of others was about equally characteristic of all groups. The responses of those who replied "Yes" to Question 7b are analyzed for Question 7c.

Responses to Question 7c were classified as "Something wrong with me" or "Other." Of those who stated that there was "Something wrong with me," nine persons said something like, "I thought maybe they were thinking I was just crazy"; the other five said, "I thought they might be wondering if I was half blind or something," or its equivalent. Members of the "Other" group either equivocated, "I was just wondering what they were thinking," or, simply accepted the fact, "I thought they probably thought I was wrong."

The responses to Question 7c provide further evidence regarding the difference between the social- and self-approval group and the self-approval group. The replies of the group high on both scales clearly resemble those of the social-approval *S*s and not at all those of the self-approval *S*s.

DISCUSSION

Figure 1 and Table 2 show that the social-approval group did, as predicted, yield to an erring majority more frequently than did groups of *S*s with other motivational profiles. The qualitative data, Questions 2a to 6a, are also in accord with the prediction. It should be emphasized that our prediction was not based on categorization of persons as conformers and nonconformers. As we have previously indicated (2, 3, 4), we assume that conformity *per se* is not a trans-situational trait, but that any individual conforms or fails to conform in a given situation depending upon his motivational profile and the relevance of conformity behavior in that situation to his motives. Thus, in a situation in which no motive gratification is to be expected from conforming, such behavior would not be predicted even where there is reason to believe social approval is an important motive in the individual's life. In an unpublished study from the authors' laboratories,[5] for example, the following interesting results were found. In an Asch judgmental situation similar to the one used in this study, greater conformity was obtained when the majority consisted of a high prestige group (college seniors) than when it was composed of peers (college freshmen). Even more striking was the finding that the presence of a low prestige majority (high school sophomores) decreased the number of yielding responses and increased the number of errors in the direction *opposite* to that taken by the majority. These findings are considered to be entirely consistent with the present interpretation of conformity behavior as instrumental activity.

The distribution of conformity scores for the social- and self-approval group shows a tendency toward bimodality, but the number

[5] This study was conducted by Misses Jane Molloy and Barbara Munger as an undergraduate course research assignment.

of cases is too small for this departure from unimodality to be significant. Such bimodality might suggest that individuals in this group choose one resolution or another of the motive conflict, i.e., they conformed or did not. The qualitative data are quite interesting in this regard.

The answers to Question 1 do little more than suggest that the social-and self-approval group is in some way different from the others. At that stage of the interview, most of the *S*s gave rather evasive answers which they contradicted later when asked about specific aspects of the experiment. For instance, five members of the self-approval group gave seemingly intropunitive responses to Question 1, but all members of that group stated that they believed they had given accurate judgments in response to Question 2a.

The data presented for question 7c in Table 3 in conjunction with the quantitative measures seem to establish the status of the social- and self-approval group. It appears that this group may have been in conflict, as we expected, a conflict which they may have resolved by differentiating between their performance and their evaluation of the majority's reaction to it. We noted the apparent lack of relationship between performance and attitude previously.

The studies presenting clear evidence of conflict in performance involve specific motor acts (7). Evidence of other conflicts, however, is typically clinical in nature (13). The implication of the literature and of our study seems to be that in the study of conflict in complex situations simple performance measures must be supplemented by other means. It appears that the social-approval *S*s yielded uncritically, or yielded but felt concern; the self-approval *S*s remained independent in performance and attitude; and the social- and self-approval *S*s did not yield, but felt concern over the majority's opinion of their independence.

The factors discussed above appear to account for the similarities and differences between the motivational groups. However, the final samples were not perfectly matched in all respects in which matching was attempted (see Table 1). The social- and self-approval group actually occupies positions intermediate between those of the social-approval group and the self-approval group on both scales 3 and 4. This fact may have contributed to the absence of more clear-cut differences between the social- and self-approval group and the others. Second, the scale 2 (security) scores of the social-approval group are higher than those of the other groups. In order to test the effect of scale 2 scores, the 41 cases were divided into those above and below the median scale 2 score for the total sample. The conformity scores of the two groups so constructed were then compared by means of the *U* test, with the finding that scale 2 scores were not related to conformity behavior in our test situation.[6]

SUMMARY AND CONCLUSIONS

Two groups of women's college students were placed in a judgmental conformity situation of the sort studied by Asch to test the hypothesis that persons with high social-approval and low self-approval motivation would yield to a unanimous, but erring, majority in judgment of length of lines more frequently than would persons with the opposite motivational profile. A third group with high motivation in both areas was included to provide information about the effects of motivational conflicts in a conformity situation. The measures of social and self approval were obtained from a four-scale measure of motivation, *The Behavior Interpretation Inventory*.

The results were in accord with the major hypothesis. The groups differed as predicted in frequency of conformity to the erroneous judgments of the majority. Comparison of the groups by pairs indicated that the high social-approval—low self-approval *S*s yielded to the majority significantly more often than did either of the other groups. The performances of the other two groups were similar.

Qualitative data obtained in postexperimental interviews indicated that *S*s in the conflict group differed from those in the self-approval group in their evaluation of the majority's attitude toward their performances. It is suggested that the high social- and self-approval *S*s experienced conflict, which they resolved by dissociating their performance from their emotional reactions to it.

[6] We do not mean to imply that concern for security (nonsocial) is necessarily unrelated to conformity behavior in all situations, but that this particular situation does not appear to have been influenced by this motive as measured.

REFERENCES

1. APPLEZWEIG, DEE G. *Scoring the open-ended form of The Behavior Interpretation Inventory.* Technical Report, ONR Contract 996 (02), NR 172-228, Connecticut College, in press.
2. APPLEZWEIG, DEE G., & APPLEZWEIG, M. H. *Stress and behavior. I. The Behavior Interpretation Inventory.* Annual Technical Report, ONR Contract Nonr 996(02), NR 172-228, Connecticut College, Aug., 1954.
3. APPLEZWEIG, M. H., & MOELLER, G. *Motivation and psychological stress: Second annual report.* ONR Contract 996 (02), NR 172-228, Connecticut College, Feb., 1956.
4. APPLEZWEIG, M. H., MOELLER, G., & BURDICK, H. Multi-motive prediction of academic success. *Psychol. Reports*, 1956, **2**, 489-496.
5. ASCH, S. E. *Social psychology.* New York: Prentice-Hall, 1952.
6. ASCH, S. E. Studies of independence and conformity: 1. A minority of one against a unanimous majority. *Psychol. Monogr.*, 1956, **70**, No. 9 (Whole No. 416).
7. HOVLAND, C. I., & SEARS, R. R. Experiments on motor conflict: I. Types of conflict and their modes of resolution. *J. exper. Psychol.*, 1938, **23**, 477-492.
8. MASLOW, A. H. "Higher" and "lower" needs. *J. Psychol.*, 1948, **25**, 433-436.
9. MOELLER, G., & APPLEZWEIG, M. H. *The Behavior Interpretation Inventory: Manual.* Technical Report No. 2, ONR Contract Nonr 996(02), NR 172-228, Connecticut College, May, 1956.
10. MURRAY, H. A. *et. al. Explorations in personality.* New York: Oxford Univer. Press, 1938.
11. OPERATIONS RESEARCH OFFICE, *et al. A study of combat stress Korea 1952 (Preliminary Report)* Tech. Memo. ORO-T-41 (FEC), Project DAS-6 FAST, Operations Res. Office, The Johns Hopkins Univer., Chevy Chase, Md., Dec., 1952.
12. ROSENZWEIG, S. The picture-association method and its application in a study of reactions to frustration. *J. Pers.*, 1945, **14**, 3-23.
13. SEARS, R. R., HOVLAND, C. I., & MILLER, N. E. Minor studies of aggression: I. Measurement of aggressive behavior. *J. Psychol.*, 1940, **9**, 275-295.
14. SIEGAL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.
15. STEIN, M. I. *The Thematic Apperception Test: an introductory manual for its clinical use with adult males.* Cambridge: Addison-Wesley Press, 1948.
16. THOMAS, R. R. *The use of structured material in projective testing.* Technical Report No. 1, ONR Contract Nonr 996(02), NR 172-228, Connecticut College, Sept., 1955.
17. THOMAS, W. I. *The unadjusted girl.* Boston: Little, Brown, 1923.

Received February 15, 1957. Prior publication.

GROUP LEARNING AND COMMUNICATION AS A FUNCTION OF TASK AND STRUCTURE "DEMANDS"[1]

JOHN T. LANZETTA

Fels Group Dynamics Center, University of Delaware

AND THORNTON B. ROBY

Institute for Applied Experimental Psychology, Tufts University

SINCE Bavelas' early papers on group structure (1, 2) a number of experiments have examined the relationships between certain of his suggested communication structure properties and group performance (e.g., 3, 4, 6, 12). In general, the experiments have employed a self-paced task with a single decision required and have utilized nets differing in the number and pattern of communication links between subjects. As Luce *et al.* have noted, "... the prototype groups are those formed by society to produce accurate information outputs in a short period of time from certain input data—information processing groups" (8, p. 82).

The present experiment is an extension of several earlier studies (7, 9, 10, 11) which, although concerned with information processing groups, focused on a different characteristic of communication structure and employed a forced-paced task. The different emphases reflected an interest in problems of division of labor in a slightly different class of information handling groups, i.e., those formed to produce *multiple* outputs from certain classes of *changing* input data, where both output and input requirements entail some division of responsibility. One critical variable manipulated when allocations of responsibilities are varied is the degree of information exchange required between group members. The term "communication structure" was used to refer to the pattern and volume of required intermember communications which resulted from the nature of the task imposed on the group and the way responsibilities were allocated within the group (9). By imposing a task with fixed, and known, display-control relationships and manipulating the allocation of observation and control action responsibilities within the group, structures were produced differing in terms of the volume and distribution of required intragroup communications.

As expected, structure difficulty decreased as the "autonomy" of group members increased, i.e., each member was the primary source of information required for his decisions, and information not directly available was located at a single external source (10, 11). That is, group efficiency decreased with an increase in the required amount of information exchange and with an increase in the amount of information exchange involving multiple sources. When the information processing demands of the task were increased, the relative differences between structures were maintained with errors for all structures increasing monotonically with increases in the amount of information presented the group per unit time. However, even under the "easiest" experimental conditions—with a low rate of information input and low information exchange structure—the groups were able to process (without error) considerable less information than might be expected from a knowledge of the information handling capacity of individuals (7).

Although the results supported the initial general hypothesis that structure-imposed demands for intragroup communication would be related in a systematic way to group performance over a range of task imposed demands, they at the same time raised additional issues. Three such interrelated issues motivated the present experiment.

1. There was some evidence that learning occurred under all of the task and structure conditions explored, but it was not possible to determine whether the learning curves were marked by sharp discontinuities,[2] and whether

[1] This research was carried out under the Air Force Personnel and Training Research Center, Lackland Airforce Base, San Antonio, Texas, in support of Project 7713. Permission is granted for reproduction, translation, publication, use and disposal in whole or in part by or for the United States Government.

[2] Group "learning" curves probably reflect the operation of two primary factors: acquisition and improvement of relevant part skills of group members and improvement in organization or "coordination." Pre-

there were qualitative and/or quantitative differences in the curves as a function of task input and structure "demands."

2. In addition, if improvement in performance occurred, it should be accompanied by changes in the communication behavior of group members. If so, would the changes in communication be a function of the requirements for information exchange imposed by structure and input variations, and could the performance of groups be related to the adequacy with which they adapted their communication behavior to these imposed "demands"?

3. Another aspect of the problem of "adaptation" was the question of the effects of group operating procedures for exchanging information on performance. As we noted above, the information processing capability of the groups was lower than might be anticipated from data on individual performance. Such a discrepancy may have resulted from the method of varying information adopted from the experiments and the task used. However, there was some indirect evidence that the primary limiting factor was the inability of groups to set up an efficient system for information exchange (7, 11).

Groups appeared to be using a "soliciting" rather than a "volunteering" system: The individual who required information called the person who was a source and requested the information rather than the source person calling the user when a change of information had occurred. Since the latter procedure entails just half the number of messages as the former, it might be expected that information transmission would be more effective using a "volunteering" system. However, volunteered information may reach a group member "out of phase" with other essential information (and would have to be stored) or at a time when other activities conflict with attending to the message (and may be lost). Thus, despite the fact that fewer messages are required, volunteering could prove to be a less effective procedure than soliciting. Data on the relative efficiency of volunteering and soliciting under a range of input and structure conditions would help clarify the issue.

The present study then, is an extension of previous experiments and is designed to throw some light on the issues raised above. Since a scarcity of data bearing on group learning or communication is available in situations of the sort considered here, no attempt is made to advance or test specific hypotheses. Rather, the focus is on obtaining objective measures that are extensive and reliable, in order to shed some light on the relationships between task and structure "demands" and group performance.

METHOD

Task

The task used in this study was essentially that described in previous reports (7, 10). There were three major components, the displays, the controls, and the communication system.

The displays were 6 AC voltmeters on each of which an aircraft "instrument" dial was simulated. The dial readings were programmed by simple resistance switching accomplished at a central console. Each instrument could be set at one of three positions. The controls were 6 two-position switches which were also labeled to simulate aircraft controls. Associated with each switch was an "operating procedure" card which indicated the correct switch settings for the various values of instrument readings. Correct settings for each switch were determined by the readings on a particular pair of instruments.

Three member groups were used. Each subject (*S*) was assigned to a three-sided booth in which was placed two simulated aircraft instruments, two control switches, and a "communication signalling" box (Fig. 1). Direct communication was prevented. The *S*s could communicate (over a standard 6-volt interphone circuit with conventional headsets and microphones) by depressing switches which provided communication connections between pairs of booths. The signalling device in conjunction with an operations recorder provided an automatic and continuous record of "who called whom" and the length of time spent in each communication. The settings of each of the six control switches were also recorded on a 20-pen Esterline-Angus operations recorder.

The task of the group was to adjust and readjust each of the six controls to their proper settings in response to periodically changing readings on the instruments. The instruments in each booth contained only part, or none, of the information required for adjusting the controls in that booth, the additional information

sumably, the form of the learning curve in any particular case will depend on the relative contribution of part skills and coordination to the group product measure. In general, however, the acquisition of part skills should produce gradual improvement in efficiency, while adoption of a new procedure or organization could produce a more marked change in efficiency. Thus, a learning curve based on group product measures may be of typical exponential form but with sharp changes in slope at various points, the discontinuities being associated with the adoption of a new procedure or reflecting a reorganization of the group. Data from the early experiments had not revealed such discontinuities, but very few trials had been used.

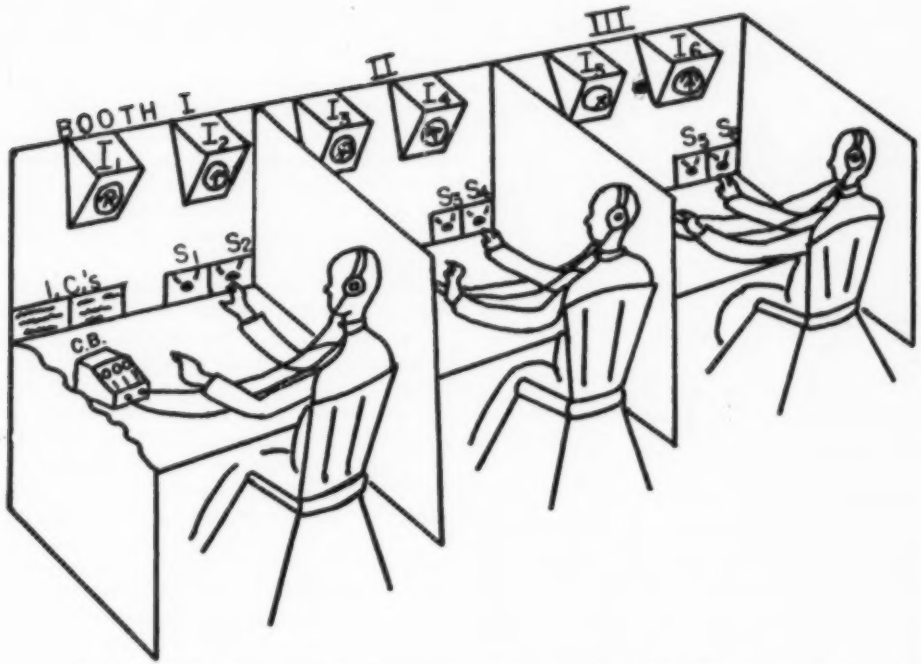

FIG 1. SKETCH OF APPARATUS AND EXPERIMENTAL ARRANGEMENT: I 1-6 = INSTRUMENTS; S 1-6 = SWITCHES; I C's = INSTRUCTION CARDS; C.B.'s = CALL BOXES

being located in one or both of the other booths. Each *S*, then, functioned as both a control agent and an information source person with responsibility for adjusting two controls and detecting and relaying instrument reading changes to other *S*s.

Experimental Conditions

Structure. Two structure conditions were employed. In the Low Autonomy Structure (LA) each *S* had two controls for which none of the four necessary instrument readings was directly available. In High Autonomy Structure (HA) each *S* had two controls for which all but one of the requisite instrument readings were available in his booth. Thus, in the LA structure, each *S* was dependent upon the other group members for all his required information, while in the HA structure each *S* was dependent upon external sources for only one reading. The interdependencies between group members represented by a matrix schema is shown in Table 1.

Task load. Task load was defined as the rate of change of the instrument readings. The input to the group consisted of the readings on each of six instruments. At a fixed interval the readings on one of the six instruments changed requiring the adjustment of either one or two controls. A *trial* consisted of 18 such changes, three for each instrument. The sequence of instrument reading changes was quasi-random, subject to the restriction that there be three changes for each instrument. The number of control adjustments that each of the three *S*s had to make in response to the 18 changes was equalized.

Three fixed intervals between instrument reading changes were employed: 6 seconds, 10 seconds, and 15 seconds. Since the number of changes was constant for a trial, this necessitated having trials of unequal length for the three different "loads."

Operating procedure. As noted above, one comparison of interest in this study was between a "volunteering" and "soliciting" procedure for information exchange. Complete control over the procedure used by the groups was unfeasible, but preliminary experience indicated that a large measure of control could be obtained by differential instructions. The initial instructions to groups assigned to the volunteering procedure stressed the role of the *S* as an information source. The *S*s were told who needed the information they had, and were asked to report any changes in their readings to the group members who required the information. In contrast, the instructions to "soliciting" groups emphasized the responsibilities of each *S* as a control agent. The *S*s were told which other group members had the information required for correctly adjusting their controls and were asked to call these *S*s to obtain the information. In both cases, several practice trials were performed under experimenters' supervision in which the appropriate procedure was again stressed.

TABLE 1
INFORMATION TRANSMISSION REQUIREMENTS IN THE TWO EXPERIMENTAL STRUCTURES*

	Group Member	Source of Information A	B	C
High Autonomy Structure (HA)				
User of information	A	3	1	0
	B	0	3	1
	C	1	0	3
Low Autonomy Structure (LA)				
User of information	A	0	1	3
	B	3	0	1
	C	1	3	0

* The entries indicate the items of information (instrument readings) held by the group member indicated in the column and required by the individual indicated in the row. Entries in the diagonal indicate items directly available.

Subjects and Design

The *Ss* were basic trainees at Lackland Air Force Base. Six *Ss* reported for testing each day; they were divided fortuitously into two teams of three *Ss* each for morning and afternoon testing. They were brought into the experimental room, seated in their booths, and oriented as to the general nature of the task and the use of the interphone equipment. Following this, they were given a brief practice session to insure that they understood how to interpret the instrument readings and the operating instruction cards. They then performed the task under one of the combinations of the experimental conditions.

There were three experimental variables: Structure, Operating Procedure, and Load, with two conditions of structure and procedure and three levels of load. A $2 \times 2 \times 3$ factorial design was used with four groups randomly assigned to each cell. Each group performed nine trials, each trial consisting of 18 instrument reading changes. After each set of three trials there was a short two- to three-minute rest period during which *Ss* remained in the booths.

RESULTS[1]

The scores used in the analysis are the number of errors in control settings made by the group within each of the nine trials. A change in an instrument reading in any of the three booths required either one or two control adjustments; if the correct control settings were not made before the next change in an instrument reading an error was recorded for each incorrect setting.

Table 2 presents the error scores for the experimental conditions and trials. The analysis of variance of these data is shown in Table 3.

Congruent with the results of earlier studies, structure and load are significant. Errors increased as the rate of change of information increased and were greater for the LA structure. The difference between the "volunteering" and "soliciting" procedures is not significant, although in the expected direction (lower errors for "volunteering").

The load $\times$ structure interaction is significant when tested against error (b) but not when tested against the significant second order interaction—load $\times$ structure $\times$ procedure. In general, the HA structure was superior to the LA structure, the difference being most pronounced for the 10-second rate of change. However, the relative superiority of HA over LA was also a function of the procedure used; a volunteering procedure resulted in a greater difference between HA and LA for the 6-second rate of change, while the converse was true for the 15-second rate of change.

Although the six conditions of structure and load cannot unequivocally be ordered in terms of difficulty, there is little doubt that the 6″-LA, and 15″-HA conditions, respectively, define the most and least difficult ends of the continuum. Thus, volunteering appears to be more effective than soliciting for intermediate difficulty levels.

Other sources of significant variance are those due to trials and the interactions of trials $\times$ load and trials $\times$ structure. Examination of the over-all means in Table 3 indicates that there was a monotonic decrease in error scores as a function of practice. Inspection of Fig. 2 and Fig. 3 indicates, however, that the regression of errors on trials was a function of both load and structure. The rate of improvement increased as the rate of change of information decreased, and was greater for the HA structure and the LA structure. There was some evidence of a performance asymptote in the 15-second trend and in the trends for the LA and HA means.

Analysis of communication records. From the 20-pen Operations Recorder record of the "calls" between group members, three interrelated measures were obtained for each trial: The number of messages initiated by all three of the group members, the average length of

[1] Acknowledgment is made for the able assistance of Mr. Edward Amberg both in experimental administration and in the statistical analysis reported.

TABLE 2
MEAN ERRORS FOR EXPERIMENTAL CONDITIONS AND TRIALS: ENTRIES ARE THE MEAN OF ERRORS FOR FOUR GROUPS

Procedure	Structure	Load	Trials									
			1	2	3	4	5	6	7	8	9	
Volunteering	LA	6″	49	49	44	47	49	46	46	52	47	429
		10″	55	38	42	41	38	39	38	34	29	354
		15″	36	33	26	20	18	20	19	18	23	213
			140	120	112	108	105	105	103	104	99	996
	HA	6″	46	42	43	37	39	34	30	23	34	328
		10″	30	21	21	11	15	13	11	10	7	139
		15″	36	25	21	17	21	20	16	16	15	187
			112	88	85	65	75	67	57	49	56	654
Soliciting	LA	6″	49	48	48	47	41	44	48	49	41	415
		10″	53	49	45	46	39	37	36	42	40	387
		15″	41	38	33	34	30	31	30	27	26	290
			143	135	136	127	110	112	114	118	107	1092
	HA	6″	47	46	40	39	39	36	35	35	35	352
		10″	36	30	23	23	17	17	16	17	15	194
		15″	23	17	15	15	10	12	11	8	9	120
			106	93	78	77	66	65	62	60	59	666

TABLE 3
SUMMARY OF ANALYSIS OF VARIANCE OF ERROR SCORES

Source of Variance	*df*	Sums of Squares	Mean Squares	*F* Ratio
Total	431	86,705.		
A. Between groups	47	64,478.	1,372.	23.66**
Load	2	28,949.	14,474.	81.31**
Structure	1	21,832.	21,832.	122.65**
Procedure	1	434.	434.	2.44
L × S	2	3,917.	1,958.	11.00**
L × P	2	454.	227.	1.27
S × P	1	253.	253.	1.42
L × S × P	2	2,245.	1,122.	6.30**
error (b)	36	6,394.	178.	
B. Within groups	384	22,227.	58.	
Trials	8	9,168.	1,146.	35.81**
T × L	16	921.	58.	1.81*
T × S	8	597.	75.	2.34*
T × P	8	370.	46.	1.44
T × L × S	16	841.	53.	1.65
T × L × P	16	589.	37.	1.16
T × S × P	8	92.	12.	.38
T × L × S × P	16	480.	30.	.94
error (w)	288	9,169.	32.	

* Significant at the .05 level.
** Significant at the .01 level.

messages, and the total time spent by all three group members in communication. It was expected that any change in communication behavior would be in terms of the frequency and duration of messages rather than in message content, since the content of messages was largely restricted to reports of instrument readings.

As with the error scores we were interested in differences in these measures of communication behavior as a function of experimental conditions, as well as changes as a function of practice. In addition, however, it was of interest to know whether there were differences between "good" and "poor" performance groups, within an experimental treatment. For the latter objective, the four groups within a treatment were dichotomized on the basis of mean error scores over the nine trials and "performance" was used as the fifth classification variable in the analyses of variance. The summary tables of the analyses of variance for number of messages, average length of messages, and total talking time, are presented in Tables 4, 5, and 6.

Number of messages. For the "number of messages" measure, load, structure, and trials contributed significant variance. The number of messages increased as the rate of change of information decreased, was greater for the LA structure, and increased monotonically with practice. There were fewer messages with the volunteering procedure than with the soliciting procedure, although the difference was not significant.

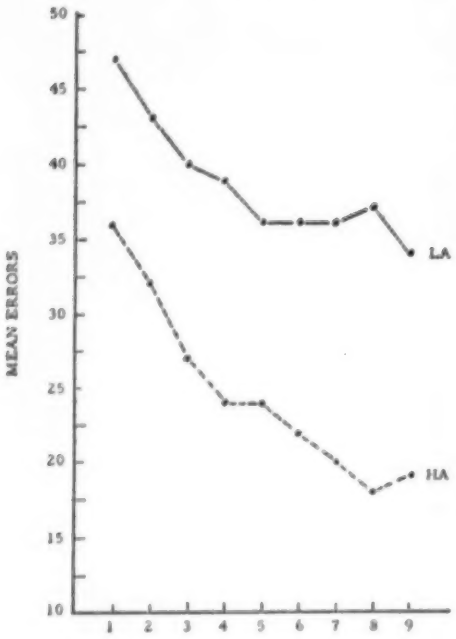

FIG. 2. MEAN GROUP ERROR SCORES OVER TRIALS FOR THE HA AND LA GROUPS

With respect to the result for the load variable, however, it was apparent that although the number of calls per trial increased as the rate of change of information decreased, the increase was proportional to the length of the trial. When corrected for length of trial, the number of messages per unit time for the three rates of change was approximately equal. The rate of message transmission appeared to be independent of the rate of change of input information; the differences between loads is attributable solely to the length of the trial.

Average length of messages. Table 5 shows that significant sources of variance for this measure are structure, performance, trials, and the trials × load × structure interaction. The average length of messages was longer for the low autonomy structure, shorter for the "better" performing groups, and decreased with practice. The changes with practice, however, were a function of both load and structure. For the LA structure the decrease was largest for the 10-second load and about equal for the 6- and 15-second loads, while for the HA structure the largest decrease was for the 15-second load with the 6-second and 10-second loads following in that order. The amount of decrease, however, was correlated with the average length of messages on trial 1 and very little difference in mean message time was evident on trial 9 or in mean message time over the nine trials. The interaction may thus be reflecting a "catching up" phenomena which was accomplished by about the fourth trial.

The surprising feature of the results was the lack of significant variance in mean message time attributable to load. Since the amount of information to be transmitted per trial was constant for the three-load conditions, one might expect that decreasing the time for communication would have decreased the average length of the communications, but this did not occur. Apparently, as the volume of information to be transmitted increased (due to structure variations) the average length of messages increased, but there was no change in average length with changes in the amount of information to be transmitted per unit time (due to load).

Total talking time. As Table 6 indicates, load,

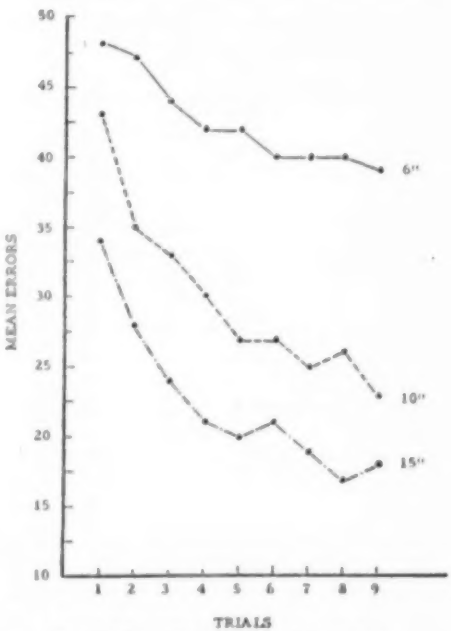

FIG. 3. MEAN GROUP ERROR SCORES OVER TRIALS FOR THE THREE LOAD CONDITIONS

structure, procedure, trials and the interactions—load × structure, and trials × structure—contributed significant variance to total talking time. Total talking time increased with a decrease in the rate of change of information, was greater for the LA structure than the HA structure, and was less for the volunteering than the soliciting procedure. Also, the difference between LA and HA structures, on this measure, increased as the rate of change of information decreased, e.g., the difference between LA and HA was greater for the 15-second load and the 6-second load.

As with the "number of messages" measure, when we corrected total talking time for the length of trials there were no differences between loads. It seems that whereas the volume of information to be relayed (structure) and the procedure used for routing information influenced the amount of time spent in communication, the rate of change of information in general, did not. However, when increased time was available for communication (decreased load), relatively more of it was utilized for sending messages in the high "demand" structure (LA) than in the low "demand" structure (HA).

The means and F ratio for trials indicate that total talking time increased with practice. The increase was greater for the LA structure than the HA structure.

TABLE 4
SUMMARY OF ANALYSIS OF VARIANCE OF "NUMBER OF COMMUNICATIONS"

Source of Variance	*df*	Sums of Squares	Mean Squares	*F* Ratio
Total	431	170,917.		
A. Between groups	47	141,898.	3,019.	
1. Load	2	68,790.	34,395.	23.33**
2. Structure	1	12,392.	12,492.	8.47**
3. Procedure	1	3,918.	3,918.	2.66**
4. Performance	1	1,306.	1,306.	—
1 × 2	2	2,314.	1,157.	—
1 × 3	2	1,888.	944.	—
1 × 4	2	2,276.	1,138.	—
2 × 3	1	5,132.	5,132.	3.48
2 × 4	1	65.	65.	—
3 × 4	1	6.	6.	—
1 × 2 × 3	2	3,340.	1,670.	1.13
1 × 2 × 4	2	906.	453.	—
1 × 3 × 4	2	1,499.	750.	—
2 × 3 × 4	1	643.	643.	—
Error (b) including quad. interaction	26	38,323.	1,474.	—
B. Within groups	384	29,019.	76.	
Trials	8	9,542.	1,193.	22.94**
T × 1	16	1,090.	68.	1.31
T × 2	8	434.	54.	1.04
T × 3	8	364.	46.	—
T × 4	8	159.	20.	—
T × 1 × 2	16	829.	52.	1.00
T × 1 × 3	16	1,092.	68.	1.31
T × 1 × 4	16	359.	22.	—
T × 2 × 3	8	805.	101.	1.94
T × 2 × 4	8	134.	17.	—
T × 3 × 4	8	394.	49.	—
Error (w) including higher interactions	264	13,817.	52.	

— *F* less than 1.
* Significant at .05 level.
** Significant at .01 level.
*** Significant at .001 level.

TABLE 5
SUMMARY OF ANALYSIS OF VARIANCE OF "AVERAGE LENGTH OF COMMUNICATION"

Source of Variation	*df*	Sums of Squares	Mean Squares	*F* Ratio
Total	431	1,709,733.	—	—
A. Between groups	47	1,364,473.	29,031.	—
1. Load	2	950.	475.	—
2. Structure	1	749,750.	749,750.	56.99***
3. Procedure	1	17,088.	17,088.	1.30
4. Performance	1	92,138.	92,138.	7.00*
1 × 2	2	14,995.	7,498.	—
1 × 3	2	31,525.	15,762.	1.20
1 × 4	2	14,541.	7,270.	—
2 × 3	1	1,576.	1,576.	—
2 × 4	1	4,300.	4,300.	—
3 × 4	1	9,343.	9,343.	—
1 × 2 × 3	2	11,389.	5,694.	—
1 × 2 × 4	2	3,837.	1,918.	—
2 × 3 × 4	1	15,588.	15,588.	1.18
1 × 3 × 4	2	55,424.	27,712.	2.11
Error (b) including quad. interaction	26	342,029.	13,155.	
B. Within groups	384	345,260.	899.	
Trials	8	73,434.	9,179.	13.66***
T × 1	16	8,312.	520.	—
T × 2	8	3,079.	385.	—
T × 3	8	9,234.	1,154.	1.72
T × 4	8	3,862.	483.	—
T × 1 × 2	16	28,980.	1,811.	2.69**
T × 1 × 3	16	16,531.	1,033.	1.54
T × 1 × 4	16	13,541.	846.	1.26
T × 2 × 3	8	1,764.	220.	—
T × 2 × 4	8	4,863.	608.	—
T × 3 × 4	8	4,289.	536.	—
Error (w) including higher interactions	264	177,371.	672.	—

— *F* less than 1.
* Significant at the .05 level.
** Significant at the .01 level.
*** Significant at the .001 level.

DISCUSSION

As might be expected, the results cast some light on the issues which gave rise to the study, but, at the same time, generate additional questions. The analysis of error scores supports the general hypothesis that "demands"

TABLE 6
SUMMARY OF ANALYSIS OF VARIANCE OF "TOTAL TALKING TIME"

Source of Variation	*df*	Sums of Squares	Mean Squares	*F* Ratio
Total	431	585,627.	—	—
A. Between groups	47	537,455.	11,435.	—
Load	2	255,420.	127,710.	87.89***
Structure	1	181,507.	181,507.	124.92***
Procedure	1	8,348.	8,348.	5.74*
Performance	1	1,846.	1,846.	1.27
1 × 2	2	39,580.	19,790.	13.62**
1 × 3	2	675.	338.	—
1 × 4	2	2,778.	1,389.	—
2 × 3	1	430.	430.	—
2 × 4	1	2,766.	2,766.	1.90
3 × 4	1	263.	263.	—
1 × 2 × 3	2	1,526.	763.	—
1 × 2 × 4	2	801.	400.	—
1 × 3 × 4	2	3,374.	1,678.	1.16
2 × 3 × 4	1	360.	360.	—
Error (b) including quad. interaction	26	37,781.	1,453.	—
B. Within groups	384	48,172.	125.	—
Trials	8	5,488.	686.	6.53**
T × 1	16	1,176.	74.	—
T × 2	8	6,072.	759.	7.22**
T × 3	8	396.	50.	—
T × 4	8	603.	75.	—
T × 1 × 2	16	1,722.	108.	1.03
T × 1 × 3	16	2,380.	149.	1.41
T × 1 × 4	16	762.	48.	—
T × 2 × 3	8	459.	57.	—
T × 2 × 4	8	385.	48.	—
T × 3 × 4	8	1,108.	138.	1.31
Error (w) including higher interactions	264	27,621.	105.	—

— *F* less than 1.
* Significant at the .05 level.
** Significant at the .01 level.
*** Significant at the .001 level.

imposed by either the group's structure or task will affect mean performance level as well as the rate of error reduction. Increasing the communication interdependency between group members and/or increasing the rate of information input, increases the average number of errors and depresses the rate of improvement with practice. In addition, over-all group performance appears to be differentially effected by the procedure used for routing information, depending upon both structure and load. A volunteering procedure is better than a soliciting system only for the intermediate levels of "demand" imposed by structure and input variations.

The chief influence of structure and task demands, however, appears to be in setting the range over which learning occurs. The form of the curves for all experimental conditions is generally quite similar; the parameters appear different. For none of the conditions is there evidence of discontinuities in the trends. Since the task appears to afford opportunity for rapid improvement in performance if optimum operating procedures are adopted, the lack of discontinuities may indicate that "insightful" reorganization did not occur, even under relatively low demand conditions.

Such an inference is in accord with the results obtained in several communication net studies. The net employed in the present study is, in essence, an all-channel or totally connected net with varying degrees of channel usage imposed by structure variations. Christie *et al.* (4) found some evidence to support a quasi-discontinuous notion of group learning in many of the nets they investigated, but little evidence of such a phenomena in a totally connected network. Guetzkow and Simon (5) demonstrated that groups in all-channel nets take relatively longer to achieve a stable and efficient organizational procedure than groups in wheel nets. Presumably as a consequence, the all-channel groups produce more continuous acquisition curves, while the wheel groups show rapid improvement in performance in the early trials and very little subsequent improvement.

The Guetzkow and Simon findings suggest that discontinuities may occur in the learning curve under conditions which facilitate the emergence of "optimum" operating procedures. In their study, the restrictions in communication which reduced the number of open channels, created an asymmetric position, and decreased the number of necessary relays, accomplished this. In the present experiment, elimination of unnecessary channels (possible in the HA structure) or restrictions on the relative availability of channels could possibly assist the development of efficient procedures, and if these procedures emerged, discontinuities in the learning curve might result.[4]

In general, the analysis of communication measures does not reveal as clear-cut effects

[4] It is possible, of course, that real progress in uncovering the factors influencing group learning will depend upon more detailed analysis of the component processes involved. In the present study error scores are composed of contributions from several "process" sources (e.g., translating instructions, detecting instrument reading changes, transmitting information) as well as several individuals, and errors from these various sources may smooth upon summation.

for the experimental conditions as does the error data. Surprisingly, the rate of change of input (load) effects no significant changes in communication behavior, although a longer total time available for communication increases the number of messages transmitted and the total time spent in communicating. Since the amount of information to be transmitted per trial is equal for all the loads it would be of interest to know what function is served by the larger number of messages under decreased load. The most plausible hypothesis is that they are redundant messages used for "checking."

Structure variation, which is coordinate with the volume of information to be transmitted, produced differences in all three measured aspects of communication behavior. In the LA structure more messages are sent, their average length is greater, and the total time spent in communicating is longer than in the HA structure. The differences are to be expected in view of the larger volume of information which had to be transmitted in the LA structure; however, the relative differences (approximately 1½:1) are not proportional to the differences in the required volume of information to be transmitted imposed by the two structures (4:1). Thus, relative to the demands imposed by the structure, the volume of transmitted information is much greater in the HA structure than in the LA structure. Since in both structures the amount of transmitted material was greater than that required by the task, it is probable that the differences reflect a much larger proportion of redundant messages in the HA structure than in the LA structure.

The difference is similar to that obtained for task load and appears to indicate that as demands decrease groups utilize the "free" time for checking activities. In view of the performance differences between load and structure conditions, this suggests that the primary determinant of performance adequacy was the degree of redundancy permitted by task and structure demands. The assumption that "free" time was employed in checking activities may also account for the lack of development of more efficient operating procedures. If, under the continued press of task requirements, all available time is devoted to activities directly task-relevant, reassessment of over-all procedure is unlikely. If true, the provision of a break in the task during which group members could freely discuss their operation may be the most general method to facilitate the development of more efficient operating procedures.

With respect to the changes in communication behavior over trials, it is seen that with increasing practice more messages are transmitted, their average length decreases, and the proportion of the trial interval used for communication increases. Apparently, groups learn to "pack" more communication into a trial interval, but the rate of learning is relatively uniform. The change in "packing" efficiency could reflect the operation of several factors: the adoption of a coding procedure, the elimination of irrelevant material from messages, or increased efficiency in performing the observation and control action functions allowing more attention to be directed to communication. Whatever the source, there is some evidence that the more effective groups "pack" more efficiently than the less effective groups (they transmit shorter messages and use a larger proportion of the trial interval for communication).

Two aspects of these results are perplexing. One of the consistent findings of the communication net studies has been a decrease in the number of messages transmitted with increasing practice. Yet, in this study, communication behavior changes have as their common consequence an increased flow of information, even though in early trials the number of messages is much larger than required by the task. Several existing differences between the net studies and the present experiment may possibly account for the discrepancy.

The net studies employ a self-paced task with information from multiple sources to be screened for a single decision. The use of message cards obviates the need for storage of the transmitted material on the part of the *S*s. The *S*'s problem, therefore, is essentially in selecting targets for messages, and as efficiency in selection increases, the number of messages required for solution should decrease. The present study uses a forced-paced task (the periodic input to the group in no way depends on antecedent decisions) with multiple information inputs and multiple required decisions.

The information is transmitted verbally and no permanent record of messa es is provided. Since every decision is dependent on collated information, a given message must reach *S* when he is free to receive it and relatively in-phase with other required information. The choice of a recipient being a trivial problem, the primary difficulty appears to be one of "phasing."

Rationally, it would seem that several procedures could reduce errors due to phasing difficulties: (*a*) advantage could be taken of predictable features of the input and the group could phase its information transmission with the predictable order of change in the information categories; (*b*) the source of information could initially transmit information once and repeat when requested; (*c*) the source could transmit repeated messages, thus allowing a recipient to sample this semicontinuous flow at his discretion.

Procedure (*a*) should lead to decreased information flow which is not consistent with the present results or with results obtained in a previous study which employed higher degrees of input predictability (7). Procedure (*b*) should result initially in greater volume of transmitted information than theoretically required although the flow should decrease with practice. Since communication differences obtained between the volunteering and soliciting procedures indicate that a strict volunteering procedure is not maintained, groups apparently did employ a procedure like (*b*). It is unlikely, however, that (*b*) alone could account for the increased information flow. Groups probably also reacted to the phasing difficulties by utilizing a technique such as (*c*).

The second perplexing aspect of the communication results is the apparent independence of the rate of change of communication behavior and the demands imposed by structure and task load. It seems justifiable to assume that environmental demands would be coordinate with drive strength and, thus, the greater the demands the more rapid the learning. Possibly, due to some process of social facilitation, the motivation for laboratory groups is higher (or lower) and less variable than is typical with individual *S*s and the influence of external factors is attenuated. At any rate, further research on the relationships between the rate of modification of important instrumental behaviors and group motivation and environmental demands is indicated.

The apparent independence of the rate of change of communication behavior and demands raises another more specific problem. Since the rate of error reduction is a function of demands, we are left in the peculiar position of trying to account for different rates of error reduction in terms of nondifferential rates of acquisition of important part-skills. One parsimonious explanation is that a given level of improvement in communication behavior has different consequences for error reduction in the various conditions. As the amount of information to be transmitted per unit time decreases, a given increase in communication efficiency could result in a proportionally greater decrease in errors. However, a precise statement of the relationship between rate of error reduction, demands of the task, and rate of improvement in communication behavior is impossible in the present study, and possibly in most studies of group learning. Methodologically, the result highlights the difficulty of accounting for group learning curves, based on traditionally employed error or time criteria, in terms of important part-processes. Similar difficulties have been encountered by other investigators (4, 8).

In conclusion, it should be pointed out that the usual strictures against overgeneralization apply quite literally in the present study. The experimental situation investigated was considerably more restricted than most group laboratory situations and the task problem was more structured than problems typically faced by either "real" groups or other temporary groups. It is interesting to note however, that despite such restrictions, much of the "explanation" of the data hinges on retrospective analysis of task requirements. Until a framework for the analysis of group task is available, hypotheses concerning the relationships between group structure and process, task demands, and group performance are going to be difficult to formulate and even more difficult to test.

SUMMARY

The present study was an extension of earlier experiments on the relationships between communication structure and group performance in an information processing task. The focus was

on changes in error performance and certain aspects of communication behavior as a function of practice under several levels of "demand" imposed by the group's communication structure and the rate of change of information input to the group. The analysis of results indicated:

1. Mean error performance and learning rate were a function of both communication structure and input load with better performance and faster rates for low demand conditions.

2. There was no evidence of discontinuities in the learning curves for any of the conditions and by inference no evidence that "insightful" reorganization occurred, even under low demand conditions.

3. All three aspects of communication behavior measured—average number of messages transmitted, average length of messages, and proportion of the trial interval spent in communication—varied with communication structure, but only the last varied with input load. There were changes in all three aspects with practice: more messages were transmitted, their average length decreased, and a greater proportion of trial time was spent in communication. Groups appear to learn to "pack" more information in a fixed interval with practice. However, there were no differences between conditions in the rate of change of communication behavior.

Apparently, left to themselves under constant situational pressures, groups modify their behavior at a characteristic rate which is independent of the demands placed on them. However, depending on the nature of the task, a given rate of behavior modification may have different consequences for the rate of improvement in some performance criterion such as errors.

REFERENCES

1. BAVELAS, A. A mathematical model for group structure. *Appl. Anthropol.*, 1948, **7**, 16–30.
2. BAVELAS, A. Communication patterns in task oriented groups. *J. acoust. soc. Amer.*, 1950, **22**, 725–730.
3. CHRISTIE, L. S. Organization and information handling in task groups. *Operations Res. Soc. of America*, 1954, **4**, 188–196.
4. CHRISTIE, L. S., LUCE, R. D., & MACY, J., JR. Communication and learning in task-oriented groups. Technical Report No. 231, Res. Lab. of Electronics, MIT. May 1952.
5. GUETZKOW, H., & SIMON, H. A. Impact of certain communication nets upon organization and performance in task oriented groups. *Management Science*, 1955, **I** (3) (4), 233–250.
6. HEISE, G. A., & MILLER, G. A. Problem-solving by small groups using various communication nets. *J. abnorm. soc. Psychol.*, 1951, **46**, 327–335.
7. LANZETTA, J. T., & ROBY, T. B. Effects of work-group structure and certain task variables on group performance. *J. abnorm. soc. Psychol.*, 1956, **53**, 307–314.
8. LUCE, R. D., MACY, J., JR., CHRISTIE, L. S., & HAY, D. H. Information flow in task-oriented groups. Technical Report No. 264, Res. Lab. of Electronics, MIT, August, 1953.
9. ROBY, T. B., & LANZETTA, J. T. Work group structure, communication, and group performance. *Sociometry*, 1956, **19** (2), 105–113.
10. ROBY, T. B., & LANZETTA, J. T. An investigation of task performance as a function of certain aspects of work-group structure. Lackland Air Force Base, San Antonio, Texas: Air Force Personnel & Train. Res. Center, June, 1956. (Research Report AFPTRC-IN-56-74)
11. ROBY, T. B., & LANZETTA, J. T. A replication study of work-group structure and task performance. Lackland Air Force Base, San Antonio, Texas: Air Force Personnel & Train. Res. Center. (Research Report in press)
12. SHAW, M. E. Some effects of unequal distribution of information upon group performance in various communication nets. *J. abnorm. soc. Psychol.*, 1954, **49**, 547–553.

Received February 25, 1957. Prior publication.

CRITIQUE AND NOTES

A FURTHER STUDY OF THE EFFECT OF STRESS ON PALMAR PRINTS

ROBERT H. DAVIS

The RAND Corporation[1]

In a recent publication, Mowrer (3) has described a method for obtaining palmar prints. Presumably densitometer measurements of such prints vary directly with emotional upset and relaxation. The terms emotional upset and relaxation are obviously difficult to tie down, but in a recent study (1) palmar prints were taken during periods which were "temporally proximal" to what are generally agreed to be stress inducing situations, oral examinations for the doctorate. Since the study reported by Beam was designed primarily to investigate the effect of real-life stress on conditioning and learning, the results with respect to palmar sweat under stress and nonstress conditions were not reported directly. It was demonstrated, however, that stress could be directly related to trials to reach a criterion of mastery, and furthermore, that the greater the increments in palmar sweating, the greater the increments in trials.

The present study was designed to test directly the hypothesis that real-life stress can be related to the index of palmar sweat. All measurements were taken during a period of stress rather than during a temporally adjacent period as in the study of Beam. Briefly, the palmar prints of a group of 34 college students were taken during an examination and contrasted with control prints of the same *S*s which were taken either two weeks prior to the examination or two weeks after the examination. The method utilized for taking the palmar prints was that described by Mowrer (3). Quantitative values were assigned to the prints by averaging five Kodak densitometer readings, one on each of the four fingers and the thumb. Rater reliability was assessed, using the rank-correlation coefficient. A separate coefficient was calculated for the experimental and control prints. Their values were .81 and .83 respectively. A nonparametric test was applied to the data. When palmar prints taken under the two conditions were contrasted using the sign test (2), densitometer values of the examination prints exceeded those of the control prints in 27 out of 34 cases ($p < .01$), lending further support to the hypothesis that palmar prints can be meaningfully related to conditions of real-life stress.

[1] The study described was conducted while the author was a member of the faculty of Allegheny College, Meadville, Pa.

REFERENCES

1. Beam, J. C. Serial learning and conditioning under real life stress. *J. abnorm. soc. Psychol.*, 1955, **51**, 543-551.
2. Dixon, W. J., & Massey, F. J. *Introduction to statistical analysis.* New York: McGraw-Hill, 1951.
3. Mowrer, O. H. *Psychotherapy: Theory and research.* New York: Ronald, 1953.

Received July 16, 1956.

SITUATIONAL DIFFERENCES IN LEADER BEHAVIOR[1]

A. PAUL HARE

Harvard University

Two types of leaders have been contrasted in the child study literature since the early 1930's, one group-oriented, the other, self-oriented (1, 6, 10, 11). In the present research, systematic observations of boys of each type are made on the school playground, where these patterns have been observed in the past, and in the more intimate, less competitive situation of the back yard.

[1] Paper read at the annual meeting of the American Sociological Society, September, 1955. The present research is part of a continuing study of the emotional development of children which has been undertaken over the last five years by members of the Harvard School of Public Health and the Human Relations Service of Wellesley. The author is indebted to Erich Lindemann, the project sponsor, for his continuing support, to Donald Klein, Elizabeth Lindemann, and Ann Ross for their collaboration in developing the research design, and to Barbara Buckstein for carrying out half the field observation. The research was completed while the author was a Public Health Research Fellow of the National Institute of Mental Health.

The definitions of group and self-oriented leaders are derived from three previous studies by Gruber (7), Lindemann and Ross (9), and Vaughan and Faber (12) who use the same population of children from which the present sample is drawn. The group-oriented leader can be described as one who leads on the basis of self-confidence, has a clear perception of the requirements of the situation, and can play a follower's role. The self-oriented leader is anxious, aggressive, and has a driving need to be in the center of the group at all times. When these general definitions are described in terms of Bales' categories (2), the group-oriented leader would be expected to be higher than the self-oriented leader in the positive and neutral acts of categories 1 through 9 and lower in the negative acts of categories 10, 11, and 12 (see Table 1).

Previous studies have not indicated the nature of the input or the average behavior of other group members directed toward the leader. However, Lindemann and Ross have observed that each leader type tends to attract followers of the same type. One would, therefore, expect the input and output profiles to be similar for each leader style.[2]

Since most of the normative studies of child behavior are based on observations made in the school setting, situational differences in behavior are not reported. Barker and Wright have noted, however, that "stable, non-psychological situations such as Sunday school, day school, etc. coerce the children to behave in relatively homogeneous ways regardless of the individual characteristics of the children" (4).

Although the observations of boys on the school playground in the present study are in part a validation of the clinical judgment of certain teachers and social scientists, the more important objective of these observations is to delineate in detail the behavior of boys in two different natural settings.

METHOD

Six group-oriented leaders and six self-oriented leaders were selected from the third grades of four New England public schools on the basis of a teacher's rating and ratings by two previous observers of the boys.[3] The teachers were first asked to identify the boys in their classes who were leaders on the playground, second, to indicate whether the boys tended to be group-oriented or self-oriented, and third, to rank the boys of each type. From the teacher's list, boys were selected who had also appeared to be group- or self-oriented in their behavior in the research of the previous investigators.

During the winter and spring of 1954 these twelve nine-year-old boys were observed on the playground during a free play period and in the neighborhood, using the set of categories for interaction process analysis developed by Bales (2). (See Table 1.) The field observation technique differed from that of Barker (3) in that (*a*) interaction was scored rather than taken down verbatim, and (*b*) individual activity apart from the group was not scored. An observer scored only one boy at a time, recording all of his acts directed to others and all acts directed to him by other group members. When possible, each of two observers made half of the observations on each subject in each situation. Although the interobserver reliability approached but did not meet the standard suggested by Bales (2) at the beginning of the research, it was well above the standard by the time the last boy was observed.[4]

The average observation period on the school playground was four sessions for a total of 47 minutes (149 acts) and in the neighborhood was four sessions for a total of 75 minutes (218 acts). Interaction with teachers, parents, or observers was not recorded; only the interaction with other children was included in the profiles. Nonverbal interaction called for by the rules of the game was also omitted.

Bales' categories were adapted for field use by removing the restriction that category 4, "Gives suggestion," *implies autonomy for the other*.

RESULTS

The average percentage of the acts in each category by leader type and situation given in Tables 1 and 2 is obtained by converting the individual percentages by the arc sine transformation, computing the means, and reconverting to percentages. In the analysis of the data, however, the arc sine values are used.

An analysis of variance for each category of acts for the *output* data (Table 1) (acts directed by the leader to other group members) indicates that the group-oriented leaders give significantly more suggestions than the self-oriented leaders.[5] They also show less tension, a difference which approaches significance.

No significant differences appear in the *input* (acts received by the leader from the other members) between the two leader types (Table 2).

[2] In previous research by the author (8), two leader styles were imposed by boy leaders leading other boys in discussion groups. Although the major distinction between the two types of leaders was the amount of participation in the discussion, the results indicate that boys with different leader styles can produce significant differences in the interaction of their peers.

[3] Elizabeth Lindemann and Ann Ross had observed these children in the school over a period of three years.

[4] When samples of mutually observed acts were compared, the reliability measure computed at the beginning of the research was 5.40 which exceeds the limit of 5.3 for 6 degrees of freedom suggested by Bales. By the end of the research the statistic was reduced to 2.74 where the maximum suggested is 6.3 for 7 degrees of freedom.

[5] The differences considered significant in this research have the probability of $< .05$ of occurring by chance. Probabilities of $< .10$ are referred to as approaching the significance level. A "one-tail" test is used only when the direction of the differences is predicted.

TABLE 1

MEAN PERCENTAGE OF INTERACTION OUTPUT FROM GROUP- AND SELF-ORIENTED BOY LEADERS FOR PLAYGROUND AND NEIGHBORHOOD*

Interaction Category	Playground		Neighborhood	
	G.-O. Ldrs.	S.-O. Ldrs.	G.-O. Ldrs.	S.-O. Ldrs.
1. Shows solidarity	5.7†	4.3	2.8	4.3
2. Shows tension release	6.4	8.5	7.8	6.9
3. Shows agreement	3.9	2.1	4.3	3.8
4. Gives suggestion	23.7	15.9	20.6	11.5
5. Gives opinion	7.2	7.3	6.5	8.0
6. Gives information	17.9	19.2	31.3	32.2
7. Asks for information	1.5	1.3	6.9	10.8
8. Asks for opinion	0.0	0.0	0.6	0.1
9. Asks for suggestion	0.0	0.0	0.1	0.0
10. Shows disagreement	1.4	3.1	4.3	3.2
11. Shows tension	4.3	10.5	3.0	4.5
12. Shows antagonism	20.1	19.5	6.1	10.5

* Average number of acts per boy is 149 on the playground and 218 in the neighborhood.

† The raw percentage is converted by the arc sine transformation for calculation and reconverted for the table. For this reason the sum of percentages for each column does not equal exactly 100.

TABLE 2

MEAN PERCENTAGES OF INTERACTION INPUT TO GROUP- AND SELF-ORIENTED BOY LEADERS FOR PLAYGROUND AND NEIGHBORHOOD*

Interaction Category	Playground		Neighborhood	
	G.-O. Ldrs.	S.-O. Ldrs.	G.-O. Ldrs.	S.-O. Ldrs.
1. Shows solidarity	8.4	8.5	5.4	3.8
2. Shows tension release	0.8	2.3	4.1	4.9
3. Shows agreement	7.8	7.8	6.3	3.4
4. Gives suggestion	12.4	10.0	16.0	14.1
5. Gives opinion	7.6	8.5	7.2	7.3
6. Gives information	8.7	14.7	32.4	32.3
7. Asks for information	2.7	2.1	10.6	8.6
8. Asks for opinion	0.0	0.0	1.1	0.2
9. Asks for suggestion	0.0	0.2	0.1	0.1
10. Shows disagreement	3.3	5.7	4.4	4.8
11. Shows tension	0.8	1.0	0.6	1.4
12. Shows antagonism	24.0	24.3	4.7	13.4

* Average number of acts per boy is 55 on the playground and 142 in the neighborhood.

Some additional differences between the two leader styles appear in only one of the two situations. The self-oriented leaders give and receive more antagonism than the group-oriented leaders in the neighborhood, a difference which approaches significance for output (Table 1) and is significant for input (Table 2). The self-oriented leaders also receive significantly less agreement in the neighborhood and receive more information on the playground (a difference which approaches significance) (Table 2).

More significant differences in output appear between the two situations for the twelve leaders (Table 1). Giving and asking for information are both greater in the neighborhood and showing antagonism is greater on the playground. Differences in giving suggestion, disagreement, and showing tension approach significance with more suggestions and tension on the playground and more disagreement in the neighborhood for both leader types.

Some of the same differences between situations also appear in the input data (Table 2). Significantly more information is given and asked for in the neighborhood and more antagonism is shown on the playground.

The consistency of the behavior of each leader is analyzed by comparing his rank order in each category (*a*) for output between situations, (*b*) for input between situations, (*c*) for output and input on the playground, and (*d*) for output and input in the neighborhood. Only 54 of the 80 correlations thus generated are positive and of these only 11 are statistically significant, i.e., greater than .78 (correlations for categories 8 and 9 were not computed). This reflects a general lack of consistency in the observed behavior, less than that reported for adult subjects by Borgatta and Bales (5).

SUMMARY AND CONCLUSIONS

Six group-oriented and six self-oriented boy leaders are observed on the school playground and in the neighborhood, using Bales' categories for interaction process analysis. The clinical judgments of teachers and social scientists are not entirely validated since boys of both leader types show the same amount of aggression on the school playground, although the self-oriented leader is more aggressive in the neighborhood. Other expected differences appear since the self-oriented leader gives fewer suggestions and shows more tension in both situations.

It is possible that the group-oriented leaders are perceived as less aggressive on the school playground because they have a more legitimate reason for aggression in the interest of group control. The self-oriented leader uses his aggression to maintain his central position so that he is usually described as "more aggressive," although both types have the same number of aggressive acts on the school playground.

The major finding of this research is that there are more differences between situations than between the two leader styles. For the twelve nine-year-old boy leaders in this sample, giving and asking for information and disagreement are higher

in the neighborhood, and giving suggestions, showing tension, and showing antagonism are higher on the playground. Differences in input are in the same direction, with more information given and asked for in the neighborhood, and more antagonism shown on the playground.

The higher amount of information in the neighborhood probably results from the fact that the children spend more time making up games, while the increase of suggestion, tension, and antagonism on the playground reflects a more competitive situation in which the individual must assert himself to maintain his status in a large group composed of children his own age.

REFERENCES

1. ANDERSON, H. H. Domination and socially integrative behavior. In R. G. Barker, J. S. Kounin, & H. F. Wright (Eds.), *Child behavior and development*. New York: McGraw-Hill, 1943. Pp. 459–483.
2. BALES, R. F. *Interaction process analysis*. Cambridge, Mass.: Addison-Wesley, 1950.
3. BARKER, R. G. *One boy's day*. New York: Harper, 1951.
4. BARKER, R. G., & WRIGHT, H. F. Psychological ecology and the problem of psychosocial development. *Child Develpm.*, 1949, **20** (3), 131–143.
5. BORGATTA, E. F., & BALES, R. F. The consistency of subject behavior and the reliability of scoring in interaction process analysis. *Amer. sociol. Rev.*, 1953, **18**, 566–569.
6. BUHLER, CHARLOTTE. Social behavior of the child. In C. A. Murchison (Ed.), *A handbook of child psychology*. Worcester, Mass.: Clark Univer. Press, 1931.
7. GRUBER, S. The concept of task orientation in the analysis of play behavior of children entering kindergarten. *Amer. J. Orthopsychiat.*, 1954, **24**, 326–335.
8. HARE, A. P. Small group discussions with participatory and supervisory leadership. *J. abnorm. soc. Psychol.*, 1953, **48**, 273-275.
9. LINDEMANN, ELIZABETH B., & ROSS, ANN. A follow-up study of a predictive test of social adaptation in preschool children. In G. Caplan (Ed.), *Emotional problems of early childhood*. New York: Basic Books, 1955, Pp. 79–93.
10. PARTEN, MILDRED B. Leadership among preschool children. *J. abnorm. soc. Psychol.*, 1933, **27**, 430–440.
11. PIGORS, P. *Leadership or domination*. Boston: Houghton-Mifflin, 1935.
12. VAUGHN, W. T. JR., & FABER, E. Field methods and techniques: The systematic observation of kindergarten children. *Hum. Organisation*, 1952, **11**, 33–36.

Received July 17, 1956.

RELIGIOUS PRACTICE AND CARDIOVASCULAR REACTIONS DURING STRESS[1]

STANLEY H. KING AND DANIEL H. FUNKENSTEIN

Graduate School of Public Health, University of Pittsburgh — *Harvard Medical School*

RECENT research has pointed to a number of psychological and sociological correlates of physiological reactions during acute stress. In terms of cardiovascular response we have demonstrated a positive association between anger directed outward and a nor-epinephrine-like cardiovascular pattern and between anger directed inward or anxiety and an epinephrine-like pattern (3). We have also shown that a subject's perception of his father as stern and dominant in discipline was associated with a nor-epinephrine-like cardiovascular reaction while perception of father as mild and nondominant in discipline was associated with an epinephrine-like pattern (6). This report describes a third correlate of cardiovascular reactions in acute stress, that of religous practice and attitudes.

METHOD

The data on which these findings are based were gathered in a two-year study of healthy college students during acute laboratory stress. The details of this study have been reported elsewhere (2, 3) but in brief the procedure was as follows. One hundred twenty-five randomly selected male students were placed in laboratory stress situations where they were given arithmetical problems to do in their heads while being hurried and criticized by the experimenter, or they were asked to repeat a story from memory while talking into an auditory feedback device. Cardiovascular measurements by means of a blood pressure cuff and a ballistocardiograph were obtained before and after each session and the *S*s were interviewed at the end of the stress situation for a measure of their emotional reactions. In addition, each *S* was seen by the *E*s on other occasions

[1] Grateful acknowledgement is given for support of this work to the School of Aviation Medicine, Randolph Field, Texas, under Contract AF 33(038)20142, The Russell Sage Foundation, the Supreme Council, 33rd Degree, Scottish Rite, Northern Masonic Jurisdiction, U. S. A., through the National Association for Mental Health, and the Laboratory of Social Relations, Harvard University.

where a battery of psychological tests and a biographical questionnaire were administered.

On the basis of relative changes between the prestress and stress measures of blood pressure and ballistocardiograph tracings the *Ss* could be divided into two main groups comparable to those described earlier by Funkenstein and Meade (4). In their study blood pressure and ballistocardiograph tracings were taken from normal *Ss* during the infusion of nor-epinephrine and during the infusion of epinephrine. The various physiological measures fell into two distinct patterns under each condition, characterized by changes in pulse, mean IJ wave on the ballistocardiograph, and mean blood pressure. The exact procedure for the determination of these two patterns can be found elsewhere (4, 5). The pattern of physiological variables in one group of *Ss* in the present study was like that found by Funkenstein and Meade during the infusion of nor-epinephrine and is thus termed nor-epinephrine-*like*. The other group had a pattern of variables comparable to that found during the infusion of epinephrine, and we have termed it epinephrine-*like*.

Data on religious practice and attitudes were obtained from the personal history questionnaire and the Religious Conventionalism scale of Levinson.

On the questionnaire each *S* was asked about the churchgoing behavior of his parents through the following questions:

1. How frequently does your father attend church? a. Frequently; b. occasionally; c. seldom.

2. How frequently does your mother attend church? a. Frequently; b. occasionally; c. seldom.

Responses to these questions were combined into a single measure of parental church attendance, as follows:

1. Regular attendance (both attend frequently or one attends frequently and the other occasionally); 2. occasional attendance (both attend occasionally or one attends frequently and the other seldom); 3. infrequent attendance (both attend seldom or one attends seldom and the other occasionally).

Further data in the area of religion were available from the scores on the Religious Conventionalism scale which had been administered as part of the battery of attitude scales developed by Levinson out of the work described in *The Authoritarian Personality* (1). Levinson has described this particular scale as follows (7):

The *Religious Conventionalism* (RC) scale contains a variety of ideas about the church, the Bible, God, prayer, and the like. High scores reflect a conception of God as punishing power figure and of the church as absolute moral authority, as well as a marked emphasis on faith, tradition, and conformity to institutional forms. Low scores, on the other hand, represent a religious humanistic or a nontheistic approach which emphasizes reason, personally derived values, and a naturalistic rather than supernaturalistic view of the world.

The *S* is presented with 12 statements in the area of religious conventionalism and asked to indicate the degree of agreement or disagreement with each statement on a seven-point scale ranging from strongly agree to strongly disagree. He can receive a score of 1–7 on each statement or a range of 12–84 on the total.

TABLE 1

PARENTAL CHURCH ATTENDANCE AND CARDIOVASCULAR REACTIONS OF SUBJECTS DURING STRESS

Parental Church Attendance	Cardiovascular Pattern of Subject		Total
	Nor-epinephrine-like	Epinephrine-like	
Regular	20	11	31
Occasional	13	19	32
Infrequent	3	12	15
Total	36	42	78

$\chi^2 = 8.719$.
$p < .02$.

TABLE 2

SCORES OF THE CARDIOVASCULAR REACTION GROUPS ON THE RELIGIOUS CONVENTIONALISM SCALE

Cardiovascular Reaction Group	Religious Conventionalism Scale		
	N	Mean	*SD*
Nor-epinephrine-like	36	42.94	19.00
Epinephrine-like	41*	33.63	18.79

* One subject became unavailable for testing after filling out the personal history questionnaire and before taking the RC scale.
$t = 2.13$.
$p < .05$.

RESULTS

The association between the measure of parental church attendance and the cardiovascular reactions of the *Ss* in stress can be seen in Table 1. The *Ss* whose parents were regular attenders were more inclined to have nor-epinephrine-like cardiovascular reactions while those whose parents were occasional or infrequent attenders had epinephrine-like patterns. This difference was especially striking between *Ss* with regularly attending parents and those with infrequently attending parents. The chi square in this table was 8.719 which with 2 degrees of freedom is significant at the .02 level of confidence.

The mean RC scale scores for the nor-epinephrine-like and epinephrine-like groups are shown in Table 2. It will be noted that even though the *SD* was large there was still a significant difference between the means with the use of a *t* test. Those with *N* patterns had higher RC scale scores than those with E patterns.

DISCUSSION

These data when combined with those on which we have reported previously indicate that there is a constellation of psychological and sociological factors which are associated with the cardiovascular reactions of healthy *Ss* in acute stress. This constellation includes the *S*'s immediate emotional

reaction, his attitudes in the area of religious values, his perception of parental behavior in discipline, and the churchgoing behavior of his parents. The *Ss* who responded to acute stress with a nor-epinephrine-like cardiovascular reaction tended to also respond with anger directed at the *E*. In addition, they were individuals who perceived their father as the principal disciplinarian in the family and stern in his discipline, they had conservative attitudes in the area of religious faith and belief, and they came from families where the church attendance habits were regular. On the other hand, those *Ss* who responded to acute stress with an epinephrine-like reaction tended to respond with anger directed against themselves or anxiety. In addition, they perceived their fathers as nondominant in discipline, had moderate to liberal attitudes in the area of religious faith and belief, and came from families where the pattern of church attendance was occasional or infrequent.

These factors come from a number of levels, not only the immediate biological (physiological and emotional), but the more enduring organization of perceptions (perception of father and attitudes toward religion), and even the surrounding family milieu (church attendance of parents).

We leave it to future research to spell out the manifold implications of these associations. We do suggest that the relationships are of sufficient strength to encourage further interdisciplinary research among the fields of physiology, psychology and sociology.

REFERENCES

1. Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. *The authoritarian personality*. New York: Harper & Bros., 1950.
2. Funkenstein, D. H., King, S. H., & Drolette, M. The experimental evocation of stress. In *Symposium on Stress*, Washington, D.C., Walter Reed Army Medical Center and Army Medical Service Graduate School, 1953, pp. 304–322.
3. Funkenstein, D. H., King, S. H., & Drolette, M. The direction of anger during a laboratory stress-inducing situation. *Psychosom. Med.*, 1954, **16**(5), 404–413.
4. Funkenstein, D. H., & Meade, L. Nor-epinephrine-like and epinephrine-like substances and the elevation of blood pressure during acute stress. *J. nerv. ment. Dis.*, 1954, **119**(5), 380–397.
5. King, S. H. Emotional and cardiovascular responses during stress. Unpublished doctor's dissertation, Harvard Univer., 1953.
6. King, S. H., & Henry, A. F. Aggression and cardiovascular reactions related to parental control over behavior. *J. abnorm. soc. Psychol.*, 1955, **50**, 206–210.
7. Levinson, D. J. The intergroup relations workshop: Its psychological aims and effects. *J. Psychol.*, 1954, **38**, 103–126.

Received June 26, 1956.

CERTAINTY OF JUDGMENT AND RESISTANCE TO SOCIAL INFLUENCE[1]

HAROLD H. KELLEY
University of Minnesota

AND THOMAS W. LAMB
Harvard Medical School

Several studies have indicated that social influences have greater effect on judgments the more ambiguous is the stimulus being judged (2, 6). One explanation of this finding is that under conditions of ambiguity, individuals are less certain about their judgments and, hence, more easily influenced. The factor of certainty has been studied independently of stimulus ambiguity by Hochbaum (4) who found high confidence in an opinion to be associated with high resistance to influence. The present note presents further evidence on the importance of certainty, obtained under conditions where there are marked individual differences in reactions to a given stimulus and associated differences in certainty of judgment.

The method involved the use of phenylthiourea (also known as phenylthiocarbamide), hereafter referred to as PTU. This substance is tasteless to some individuals but tastes extremely bitter to others. After preliminary work indicated that we could present PTU in a fairly controlled manner and readily distinguish tasters from nontasters,[2]

[1] Conducted while the authors were at Yale University, under the auspices of the Yale Communication Research program. The authors are indebted to several chemists at Yale and Minnesota for their advice.

[2] Blakeslee (1) has found that tasting behaves as a dominant gene and nontasting as a recessive gene. The basis for this difference is now thought to exist in the

we used it to study the following question: Does the effect of a majority depend upon the particular opinion they advocate, relative to the opinion held by the minority? For example, will the effect of the majority's espousal of "A" when the minority initially favors "B" be the same as if the majority had supported "B" and the minority, "A"? That this problem readily lends itself to the use of PTU is apparent: We have merely to form groups composed of unequal numbers of tasters and nontasters, in some of which the tasters are in the majority and in others of which the nontasters predominate.

METHOD

A rating form was prepared with three scales of "pleasantness of taste," ranging from 0 (bad) to 10 (good). The scales were labeled A, B, and C, respectively, to correspond to three gummed labels, also labeled A, B, and C, which were attached to the bottom of the forms. Label A was a plain label, and Label B was the same kind of label coated with pineapple flavor. Label C was prepared with a combined coating of PTU and cherry flavoring. Scales B and C each had an accompanying scale for degree of certainty, ranging from 1 (not at all certain) to 7 (very certain).

The subjects (*Ss*), Yale University upperclassmen, were first seen individually and privately. Each was told that we were trying to find out "whether there are absolute values for these tastes or whether there are individual differences in ratings of them." He was then given a form, asked to taste Label A, and instructed to give it a rating of 5 (average) on the scale of pleasantness. He was then asked to taste B and C, and to rate each one in relation to A. After rating B, he was asked to rate his certainty that B should be given the pleasantness rating he had given it. A similar procedure was followed for C.

On the basis of these private judgments, the *Ss* were scheduled in 30 three-man groups, half composed of one taster and two nontasters and half, of two tasters and one nontaster. Because the population of *Ss* yielded a ratio of tasters to nontasters of approximately 3:1, about 150 individuals were privately tested before the 30 groups could be arranged. In scheduling each group, the further requirement was made that each of the three men should know the other two, the result being that men who lived near each other in a given residential college were placed in the same group.

The groups were assembled one to three days after the individual testing. The instructions were similar to those used earlier, except that certainty ratings were not requested and the *Ss* were asked to announce their ratings publicly, each one first giving his rating of B, and then of C. Each time, the majority persons, whether tasters or nontasters, were asked to announce their ratings first. When it came his turn on B, the minority person usually found that the others had given ratings similar to his private ones. But on C, he found that their ratings were markedly different from his own evaluation.

RESULTS

The data were analyzed only for the 90 *Ss* used in the group portion of the experiment.

1. The tasters and nontasters agreed very closely in their ratings of Label B, the one coated with pineapple flavoring. On the average, they gave it a rating near 7.0 on both the initial private ratings and the publicly announced ones. We mention this to show that the tasters and nontasters are comparable in their reactions to a substance not containing PTU and that they used the rating scale in much the same manner.

2. The tasters' initial private ratings of C, the experimental label coated with PTU and cherry flavoring, were markedly lower than the ratings made by the nontasters. The mean rating for the 45 tasters was 0.8 and for the 45 nontasters, 6.7. A t test of this difference yields a value of 32.7 which is obviously significant. The two distributions of ratings do not overlap at all. The range of ratings given by the tasters is from 0 to 2; for the

---

interference of PTU with the activity of certain enzymes (7, Volume I, p. 337). It is also thought to be a goiterogenic compound if administered on a long-term basis, as it inhibits the formation of thyroxine, a hormone of the thyroid gland (7, Volume II, p. 427). It has been used as a food preservative, but recently in this connection the question of its possible carcinogenic action has been raised (5).

PTU can be obtained in a pure form from the Eastman Kodak Company or in a semipure form from many chemical supply houses. It can be recrystallized from alcohol, since it is extremely soluble in hot alcohol and relatively insoluble in cold alcohol. A negative cyanide test should be obtained before the chemical is used on *Ss* (3, p. 258).

Because of the possible dangers of large quantities of the compound and the nauseating effects on some persons, it is desirable to present PTU in very small quantities. At the same time, trial taste-tests indicated that strong concentrations should be used if an immediate, relatively standard reaction is desired. With weak concentrations many tasters have a sensation of bitterness only as an "aftertaste" which occurs up to ten seconds after the presentation. These considerations were met by presenting PTU in the form of a dried solution on the back of ordinary gummed labels. This method provides the *S* with a high concentration of PTU, but at the same time insures that he ingests only a minute quantity of it.

The solution was prepared by diluting to twice its volume a saturated solution of PTU in Baker Extract Company's "True Fruit Cherry Flavor" which has a 16% alcohol base. It was then applied with an atomizer to the back of Reyburn's label No. 13 until a deep purple color was attained. This rather crude method of "dosage" was found to produce quite standard results, because of the high concentration of PTU that was used. Individual differences in reaction were noted, but of the *Ss* to whom we presented the labels, not one exhibited a violent reaction, yet all the "tasters" experienced an intense, bitter taste.

nontasters, from 4 to 8, with 42 of the 45 giving ratings from 6 to 8.

3. The public judgments of Label C were first examined for the *Ss* in each of the 30 groups who announced their ratings first. These ratings, when compared with the private ones made by the same 30 persons several days earlier, indicate the stability of these taste judgments in the absence of experienced disagreement. The scatterplot of private versus public (early vs. late) ratings reveals two separate clusters of *Ss*, the tasters and the nontasters. The product-moment correlation between the early and late ratings is +.98. Put in other terms, the data showed that 19 of the 30 *Ss* gave identical ratings on both occasions, 10 gave ratings which differ by one point, and 1 gave ratings differing by two points. The results were virtually identical for the *Ss* (also members of the majorities) who announced their ratings second. It is quite clear, then, that in the absence of pressure to change, the private ratings were highly predictive of the later public ones.

4. The public ratings of C were next examined with respect to the changes from the earlier private ratings exhibited by the *Ss* who announced their judgments last. Since these *Ss* were always in the minority, these changes indicate the effects of the majority opinions. There is a marked difference between the minorities composed of tasters and those composed of nontasters. The average amount of change for the tasters was 0.0, whereas the average for the nontasters was −2.6, the negative sign indicating that they made a lower (less "pleasant") rating in the group situation. There is clearly a significant difference between the two kinds of *Ss* in their tendency to go along with the majority opinion. Only 4 of the 15 tasters shifted in the indicated direction (toward a higher rating) while 13 of the 15 nontasters shifted toward a lower rating (with Yates' correction, chi square = 8.7, $p = .004$). It appears, then, that a taster majority has a greater effect on a nontaster minority than a nontaster majority has on a taster minority.

Two possible bases for this difference are suggested by our data, only the second of which seems to be important. First, it happened that the average discrepancy between the ratings announced by the majority members and the private rating previously given by the minority member was somewhat larger in the groups where the tasters were in the minority. While this difference between the two kinds of groups approaches significance (p = approx. .06), we doubt that it accounts for the observed effect. It seems unlikely that the minority tasters would have changed less simply because the majority members' opinions were somewhat farther away from their private ones. In fact, within each sample, there seems to be the opposite tendency: greater change tends to be associated with a larger discrepancy. If we compare *Ss* for whom the discrepancy was identical in size, there is still a marked tendency for nontasters to exhibit more change than tasters.

Second, tasters and nontasters were found to differ in how certain they were about their ratings of the pleasantness of Label C. The average certainty rating for the tasters was significantly higher than that for the nontasters (6.6 as compared with 5.4; $t = 6.25$, $p < .001$). This difference was also significant, at the .01 level, when only the *Ss* used as minority members were compared. This result suggests that the reason for the tasters' great resistance to majority influence may be found in the greater intensity of their opinions about PTU. The correlation between change and certainty within each sample is equivocal on this point because the range of both variables is severely restricted when we consider only the tasters or nontasters. We may note parenthetically that for both Labels B and C, certainty shows the typical U-shaped relation to extremity of rating on pleasantness. The closer a person rated the flavor to the extreme ends of the scale, the more certain he was of the correctness of his rating. The eta coefficient is highly significant in both cases.

SUMMARY

In a simple social influence problem involving judgments of taste with PTU the critical stimulus, it was found that the effects of majority opinion were markedly different depending upon whether the minority persons were tasters or nontasters of PTU. The data suggest that this asymmetric effect, whereby nontasters are more susceptible to majority influence, may be attributable to the stronger reactions tasters have to PTU and the resulting greater certainty they have about their judgments of the substance.

REFERENCES

1. BLAKESLEE, A. L. Genetics of sensory thresholds: Taste for phenythiocarbamide. *Proc. nat. Acad. Sci.*, 1932, **18**, 120–130.
2. COFFIN, T. E. Some conditions of suggestion and suggestibility. *Psychol. Monogr.*, 1941, **53**, No. 4.
3. FEIGL, F. *Spot tests. Vol. I. Inorganic applications.* Houston: Elsevier, 1954.
4. HOCHBAUM, G. M. The relation between group members' self-confidence and their reactions to group pressures to uniformity. *Amer. sociol. Rev.*, 1954, **19**, 678–687.
5. LEAR, J. Food and cancer. *Saturday Rev.*, 1956, **39**, No. 40, 57–62.
6. LUCHINS, A. S. Social influences on perception of complex drawings. *J. soc. Psychol.*, 1945, **21**, 257–263.
7. SUMNER, J. B., & MYRBACK, K. (Eds.), *The enzymes.* New York: Academic Press, 1950–1952.

Received August 10, 1956.

TWO KINDS OF ASSUMED SIMILARITY BETWEEN OPPOSITES[1]

IVAN D. STEINER AND COLIN G. McDIARMID

University of Illinois

In a series of recent studies, F. E. Fiedler (3, 4, 5) has shown that a leader's "assumed similarity between opposites" (ASo) is an important correlate of his team's efficiency. Basketball teams and surveying teams were shown to be more efficient when the sociometrically most-chosen member attributed very different psychological traits to persons he regarded as good and bad co-workers. The efficiency of bomber and tank crews was also found to be related to the leader's ASo. This paper reports an investigation of some statistical and psychological properties of ASo.

ASo is a measure of the magnitude of the differences between two profiles. One profile represents the individual's ratings of a good co-worker on twenty or more six-step graphic rating scales. The ends of each scale are identified by polar terms such as "energetic" and "tired." The second profile consists of the individual's ratings of a poor co-worker on the same graphic rating scales. The ASo score is obtained by summing the squares of the distances separating these two profiles from one another (i.e., the distances separating the ratings given to the good co-worker from the ratings given to the bad co-worker). Thus:

$$\text{ASo} = \sum_{i=1}^{N} (g - b)^2$$

when $(g - b)$ equals the discrepancy between the ratings given to the good and the bad co-workers on one of the graphic rating scales, and N equals the number of rating scales employed. The ASo score may be broken into two component parts (2):

$$\frac{\text{ASo}}{N} = (M_g - M_b)^2 + \sigma^2_{g-b}$$

where M is the mean of the several ratings given to a co-worker. The $(M_g - M_b)^2$ component is the square of the average *elevation difference* between the two profiles—the tendency of one profile to lie above the other. Intuitively, we can view this component as reflecting the individual's tendency to regard one co-worker as endowed with more desirable attributes than the other. The second component measures the degree to which the two profiles have different shapes, regardless of their average elevations. This component may be viewed as reflecting the individual's tendency to regard his two co-workers as possessing *different combinations* of good and bad traits.

There is no logical necessity for these two components of the ASo score to be correlated. It is entirely possible for a small or zero average elevation difference to occur along with a large difference in profile shapes. And a large average elevation difference could occur with a small difference in profile shapes. The ASo score is, therefore, a gross measure which combines two logically independent measures. The purpose of this research is to determine whether the two components of ASo are also empirically independent of one another, and, if so, to determine whether the two components correlate differently with other measures. The other measures employed in this research include scores on the California F scale (1), and scores obtained from self-ratings. Reasons for expecting scores on the F scale to correlate with one or both components of ASo will be discussed later.

METHOD

Forty-two members of an undergraduate class in psychology served as subjects (*S*s) for this study, but because of omissions from the data, the *N*s involved in the reported results range from 38 to 42. The two sexes were approximately equally represented in the usable samples.

ASo Scores

*S*s responded to Fiedler's instrument consisting of 20 graphic rating scales, the ends of which were identified by polar terms. As is customary with this instrument, *S*s rated themselves, a good co-worker, and a bad co-worker on each of the 20 scales. The *S*s were asked to think of someone with whom they had been able to work very well on some occasion in the past, and of someone with whom they had had difficulty working. Each *S*'s ASo score was computed by obtaining the sum of the squared distances separating his ratings of these two co-workers.

Elevation Difference Scores

In order to compute the $(M_g - M_b)$ component of ASo it was necessary to obtain judgments concerning which end of each scale was regarded by *S* as the good end. For this purpose each *S* rated himself on Fiedler's instrument two additional times, once giving himself the benefit of any doubts whatever, and once without giving himself the benefit of any doubts. Inspection of these two sets of self-ratings permitted a determination of the good-bad direction of the scale for the *S*. Ratings at the extreme bad end of a scale were then assigned a value of one and those at the good end were assigned a value of 6. The mean value of the ratings given to the good co-worker was obtained, as was the mean value of ratings given to the bad co-worker. The difference between these two means is the *S*'s Elevation Difference score.

[1] This research was partially supported by a grant from the University of Illinois Research Board.

Profile Dissimilarity Scores

Each *S*'s Profile Dissimilarity score was obtained by subtracting his squared Elevation Difference score from his $\frac{\text{ASo}}{N}$ score. The Profile Dissimilarity score is the square root of the difference obtained in this manner.

Self-Ratings

As noted above, each *S* rated himself three times: once attempting to be as accurate as possible (realistic self-ratings), once giving himself the benefit of any doubts (optimistic self-ratings), and once without giving himself the benefit of any doubts (pessimistic self-ratings). The mean and standard deviation of these three sets of self-ratings were computed for each *S*.

F Scale. A month after the previous data had been obtained each *S* responded to the F scale. Items were scored in the standard manner.[2]

RESULTS

As should be expected, both Elevation Difference scores and Profile Dissimilarity scores were found to be positively correlated with ASo scores. The product-moment correlations are .75 and .53 respectively. With a sample of 42 persons, both correlations are significantly different from zero at the .01 level. The correlation between Elevation Difference scores and Profile Dissimilarity scores is −.03. The two components of ASo appear to be empirically independent of one another.

Relationships between each of the three measures and other variables are indicated by the product-moment correlations shown in Table 1. Three different patterns of relationships are discernible. F scores correlate significantly with ASo scores, but correlations with the two components of ASo are low and do not reach statistical significance. Standard deviations of self-ratings do not correlate significantly with either ASo scores or Elevation Difference scores, but do correlate significantly with Profile Dissimilarity scores. Finally, the mean of the individual's optimistic self-ratings (i.e., the mean of the self-ratings made by *S* when he was instructed to give himself the benefit of any doubts) fails to correlate significantly with either ASo scores or Profile Dissimilarity scores, but does correlate significantly with Elevation Difference scores. (Neither the mean "realistic" self-rating scores nor the mean "pessimistic" self-rating scores correlated significantly with ASo or either of its components.)

These findings thus indicate that the two components of ASo sometimes correlate in different degrees (and possibly in different directions) with other variables. The components may also combine to produce ASo scores which correlate more highly with other variables than do either of the components taken separately.

[2] Form 45 of the F scale (1, pp. 255–257) was employed in this research, but item number 22 was deleted because it had become obsolete since the time the scale was published.

TABLE 1

CORRELATION OF ASo AND ITS TWO COMPONENTS TO OTHER VARIABLES

	F Scale Scores	Mean of Self-ratings	Sigma of Self-ratings	Mean of Optimistic Self-ratings
	N			
	38	42	42	42
ASo scores	−.43**	.02	.29	.21
Elevation Difference Scores	−.21	.18	−.07	.35*
Profile Dissimilarity Scores	−.29	−.09	.62**	−.10

* Reaches the .05 level.
** Reaches the .01 level.

DISCUSSION

The ASo score, according to these results, is a gross measure that combines two uncorrelated variables. One of these variables is the perceived discrepancy between the overall "goodness" of the traits possessed by two co-workers; the other variable is the perceived dissimilarity between the patterns of traits possessed by two co-workers. Each of these two variables contributes substantially to the ASo score, but two *S*s may obtain the same ASo score even though their scores on the two components of ASo are very different. For example, two *S*s in the present study obtained ASo scores of 166; for one *S* the $(M_g - M_b)^2$ component was responsible for seven per cent of the ASo score, whereas for the other *S* this component was responsible for 82 per cent. For most purposes it would probably be unrealistic to treat these two persons as though their ASo scores were equivalent.

When ASo scores are related to an external variable, a significant correlation may be obtained because either or both of the components of ASo are correlated with the external variable. Or, an insignificant correlation may be obtained because neither component is correlated with the external variable, or because both are correlated but in opposite directions. Since it is quite possible for one or both components to correlate much more highly with an external variable than does ASo itself, it is recommended that studies of "assumed similarity between opposites" deal with the components as well as with ASo.

In this research the direction of the relationships between F scale scores and the two components of ASo was not predicted in advance, but significant relationships were expected. According

to Adorno et al. (1), the high F scorer is reluctant to see differences between members of his in-groups; all members are good, whereas people in out-groups are bad. On the other hand, the low F scorer is described as willing to see various combinations of traits in members of his in-group, and to minimize differences between members of in-groups and out-groups. Consequently, if one assumes that both the good and the bad co-workers are regarded by the *S* as members of his in-group, both components of ASo should be small for high F scorers and large for low F scorers. But, if the poor co-worker is regarded as a member of an out-group, both components of ASo should be larger for the high F scorer than for the low F scorer. The findings reported in Table 1 suggest that when people are asked to think of someone with whom they have worked poorly, they tend to think of an in-group member.

There are two possible interpretations of the .62 correlation between Profile Dissimilarity scores and the standard deviations of *Ss*' realistic self-ratings. Perhaps a response set leads some people to give ratings that are highly variable from item to item, and leads other people to give ratings that are rather constant from item to item. If these response sets operated both when *S* rated himself and when he rated his good and bad co-workers, a correlation between Profile Dissimilarity scores and the standard deviation of realistic self-ratings might result from the response set alone. However, if such response sets are operating, they should be expected to affect *Ss*' optimistic and pessimistic self-ratings just as they affect ratings of good and bad co-workers. Inspection of the data shows that when standard deviations of *Ss*' optimistic and pessimistic self-ratings are computed and added to the standard deviations of their realistic self-ratings, the correlation between the sums of these three standard deviations and the Profile Dissimilarity scores falls to .20. This finding tends to contradict an explanation based on the assumption of a broad, "mechanical" response set, and suggests a second explanation. Possibly some persons really do see themselves and other people as the possessors of a mixture of good and bad traits, whereas other persons tend to see both themselves and their associates as the possessors of rather uniformly good or uniformly bad traits. The data of this research do not provide a firm basis for choice between these two explanations. However, whatever the proper explanation may be, it is clear that Profile Dissimilarity scores represent a variable and potentially significant component of ASo, and that this component is empirically independent of the Elevation Difference component.

The findings of this study suggest that Fiedler's highly productive research might be made even more valuable if ASo scores were broken into their component parts. Such a step might also be helpful in clarifying the reasons why the ASo of a leader is related to his group's efficiency. Fiedler's explanation of his results rests upon assumptions concerning the personalities of high and low ASo leaders. For example, the leader with a large ASo score is said to be "critical, reserved, and rejecting, and especially so toward men who he considers to be poor co-workers." He is presumed to discriminate sharply between good and bad co-workers, to resist emotional involvement, and to be guided by task-oriented attitudes. These assumptions concerning the personal qualities of leaders with large ASo scores have not yet been tested, and Fiedler himself has indicated that he regards them as tentative. In light of the present study it is reasonable to guess that leaders with identical ASo scores may have quite different personal qualities depending upon the size of the components of their ASo scores. Somewhat more complex assumptions concerning the relationships between ASo and personality variables would seem to be needed.

SUMMARY

The conventional measure of assumed similarity between opposites (ASo) can be broken into two components which are logically and empirically independent of one another. Taken separately these two components sometimes correlate more highly with external variables than does ASo itself. Implications of these qualities of ASo are discussed.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D.J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper & Bros., 1950.
2. CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.
3. FIEDLER, F. E. *Assumed similarity measures as predictors of team effectiveness in surveying.* Urbana, Ill.: Bureau of Research & Service, Univer. of Illinois, 1953. (Mimeo.) (Tech. Rep. No. 6, Contract N6ori-07135.)
4. FIEDLER, F. E. The influence of leader-keyman relations on combat crew effectiveness. *J. abnorm. soc. Psychol.*, 1955, **51**, 227–235.
5. FIEDLER, F. E., HARTMANN, W., & RUDIN, S. A. *Correction and extension of the relationship of interpersonal perception to effectiveness in basketball teams.* Urbana, Ill.: Bureau of Research & Service, Univer. of Illinois, 1953. (Mimeo.) (Tech. Rep. No. 5, Contract N6ori-07135.)

Received June 27, 1956.

CASTRATION, CIRCUMCISION, AND ANTI-SEMITISM

IRVING E. ALEXANDER AND SHELDON BLACKMAN[1]

Princeton University

In his writings, Freud gave recurrent expression to the view that fear of castration is a causal factor in anti-Semitism. In 1909 he wrote:

> The castration complex is the deepest unconscious root of anti-Semitism; for even in the nursery little boys hear that a Jew has something cut off his penis—a piece of his penis, they think—and this gives them a right to despise Jews (4, p. 179).

In later statements he pointed out that castration and circumcision are equivalent in the unconscious (5), and also that the strangeness of the custom for gentiles serves to activate unconscious castration fear (6). That these views found some measure of acceptance in psychoanalytic circles can be seen in the work of Fenichel (2).

As a consequence of changing medical attitudes toward infant care, circumcision is no longer confined to one group in our culture. This turn of events provides an opportunity to gather some controlled information pertinent to Freud's speculation. The particular question we shall ask is whether circumcision is a relevant variable in anti-Semitism as measured in a contemporary male gentile population.

Method

The measure of anti-Semitism used in this study is the 52-item A-S scale described by Levinson and Sanford (8). This scale was chosen because it has been standardized on a college population and some evidence is available for its reliability and validity. A further point in its favor for the present purpose is that it includes a "threatening" scale consisting of ten items related to Jews as powerful and possible overpowering figures. In the case of all the test items, a score of seven indicates strong agreement with the anti-Semitic position, a score of one, strong disagreement. If, as happened rarely, an item was omitted or for other reasons could not be scored, a value of four was arbitrarily assigned. Appended to the test was a data sheet on which relevant background information was recorded. Included was a section on health which contained questions on surgery undergone, e.g. appendectomy, tonsillectomy, circumcision, etc., and the time at which the operations had been performed.

This scale was administered to 381 students drawn from various psychology courses at Princeton and Rutgers Universities.[2] The subjects (*Ss*) included in this study are 142 gentile male undergraduates selected from the larger population and divided into two matched groups of 71 each. The groups were matched on the basis of background variables the relationship of which to anti-Semitism had been investigated in the literature (7). These variables include family income, geographical location of home town, religion, church attendance, and frequency and intimacy of contact with Jews. The sample is relatively homogeneous with regard to other background variables such as age, rural-urban differences, and intelligence so that no effort other than inspection was made to control these factors. The one variable upon which the groups are known to differ systematically is whether or not the members are circumcised.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF SCORES ON THE A-S SCALE

Group	Mean	*SD*
Full Scale		
Circumcised	146.5	42.8
Uncircumcised	151.8	43.4
"Threatening" Scale		
Circumcised	26.4	7.5
Uncircumcised	27.4	9.4

Results

Means and standard deviations for each group on both the full scale and the "threatening" subscale are presented in Table 1. In both instances, on the full scale and on the threatening scale, the group mean for the uncircumcised *Ss* exceeded that of the circumcised. However, these differences were small and statistically unreliable. An examination of the *F* ratios (full scale $F = .43$, threatening scale $F = 1.01$) gives no grounds for rejecting the view that they are attributable to chance.[3] The results of this study do not support the hypothesis that circumcision is a relevant variable in determining attitudes towards Jews and in turn cast doubt upon Freud's assumption as to the relation of castration anxiety to anti-Semitism.

Discussion

Several considerations come to mind in assessing the bearing that this study might have on the

[1] Now at the Ohio State University.

[2] We are grateful to Mr. Edward Nolan of Rutgers University, Newark Branch, for assistance in gathering the sample.

[3] The analysis of variance technique does not take into account the positive correlation introduced by the matching of the groups. No satisfactory technique exists for estimating the correlation when groups are matched on the basis of total composition. The magnitude of F in each case is so small that the correlation would have to be above .9 in order to change the interpretation of our results.

castration–anti-Semitism hypothesis. One concerns the apparent evolution of attitudes toward this practice in the last twenty years. If few people in the culture at the present time, circumcised or not, have negative feelings about infantile circumcision, one might argue that this change has obscured the validity of Freud's original contention. The custom is no longer strange, and therefore ought not to arouse castration fear. This argument we would grant for children born in the present and future decades, but not for children born during the early stages of cultural change in attitude. During our *Ss*' formative years, both positive and negative attitudes toward circumcision were widespread and accepted in the larger culture.

Another consideration is whether we were dealing with a subculture so demanding of conformity from its members (the sample was predominantly composed of Princeton students) that it tended to obscure any real underlying attitudinal differences. That is to say, perhaps we were getting a reflection of a Princeton stereotype towards Jews rather than a true indication of the *S*'s feelings. The fact that the scores of Jewish and gentile students in the entire Princeton sample differ significantly ($p = .05$) suggests that the demands of the subculture leave room for expression of attitudinal differences.

Recently, two other investigators have presented discussions on the relationship of circumcision to castration which bear indirectly on the question we are considering. Bettelheim (1) and Fraenkel (3) have both questioned the widespread psychoanalytic belief that circumcision as a ritual stems from or has the meaning of a castrating act. Bettelheim, after a study of some anthropological and clinical material, finds no general evidence to support the idea that the practice arose in order to maintain an incest taboo or even that it evolved from father dominated societies. The Jewish practice, a relatively late one in history, he asserts is atypical in that it is carried out in infancy and is restricted entirely to males. The meaning of the circumcision rites, according to Bettelheim, is more frequently related to fertility and to the individual fulfillment of an unconscious bisexual wish. He offers no support for the assumption that circumcision and castration are necessarily symbolically equivalent or that circumcision would inevitably have negative valence in the unconscious.

Fraenkel concluded from a study of ancient Jewish religious treatises, particularly the Pentateuch, that circumcision appeared to be more related to the benediction of fecundity than to an attenuated castration inflicted by the chief on his son. Only after circumcision was Abraham able to beget Isaac. Fraenkel characterizes the promise of the ritual by a quotation from the book of Genesis, "His seed will be as numerous as the sands of the earth." If the circumcision-castration argument cannot therefore be invoked with confidence, then Freud's speculation with regard to anti-Semitism loses its *raison d'être*.

SUMMARY

To obtain data relevant to Freud's speculations on the relationship between circumcision and anti-Semitism, attitudes towards Jews were measured among matched groups of circumcised and uncircumcised male gentile college students. No reliable differences in attitude were found between circumcised and uncircumcised subjects. These findings indicate the relative unimportance of circumcision as an isolated factor in anti-Semitism. The implications for the relationships involving circumcision, castration, and anti-Semitism are discussed.

REFERENCES

1. BETTELHEIM, B. *Symbolic wounds*. Glencoe, Illinois: Free Press, 1954.
2. FENICHEL, O. Elements of a psychoanalytic theory of anti-Semitism. In E. Simmel (Ed.), *Anti-Semitism*. New York: International Universities Press, 1946.
3. FRAENKEL, E. La circoncision chez les Juifs peut-elle s'expliquer comme une castration atténuée infligée à ses fils par le chef de la horde? *Psyché*, 1952, **7**, 377–385.
4. FREUD, S. *Collected papers*. Vol. 3. An analysis of a phobia in a five year old boy. London: Hogarth Press, 1925.
5. FREUD, S. Eine Kindheitserrinerung des Leonardo Da Vince. *Gesammelte werke*. Achter Band. London: Imago Publishing Company, 1943.
6. FREUD, S. *Moses and monotheism*. New York: Vintage Books, 1955.
7. HARDING, J., KUTNER, B., PROSHANSKY, H., & CHEIN, I. Prejudice and ethnic relations. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. II. Cambridge: Addison-Wesley, 1954. Pp. 1021–1061.
8. LEVINSON, D. J., & SANFORD, R. N. A scale for the measurement of anti-Semitism. *J. Psychol.*, 1944, **17**, 339–370.

Received September 13, 1956.

CONSISTENCY IN RESPONSE TO GROUP PRESSURES[1]

STANLEY ROSNER

Society for Mental Hygiene, Bridgeport, Connecticut

THE studies to be described take as their point of departure the investigations of Asch (1, 2) on modification of judgments under group pressure in face-to-face situations. Asch has reported that naive subjects (*Ss*) yielded approximately one-third of the time to the erroneous judgments of a tutored majority in matching the length of comparison lines with a standard line. The *Ss* who yielded early in the line series (consisting of 12 critical trials) tended to continue to yield throughout, while *Ss* who were independent of the announced judgments of the tutored majority, tended to continue independent throughout.

The purpose of the present studies is to answer the question: Are the consistencies in yielding or not yielding specific to the situation studied, i.e., judgment of lines, or do they extend beyond it? Is such consistency episodic or situationally determined, or is it a consequence of relatively settled modes of orientation?

METHOD

To answer the above question, 88 student nurses[2] were observed in each of several Asch-type situations, in each of which they had to make a choice between being independent or conforming to the judgments of three accomplices[3] who were tutored by the experimenter to make erroneous judgments.

Tasks[4]

Three memory tasks (nonsense words, nonsense figures, and a simple paragraph), a questionnaire, and the line series used by Asch (2) were employed in the experiment. Necessary controls for order of presentation and equating of control and experimental series were observed.

RESULTS

1. Where comparisons are possible, present findings confirm those of Asch (1, 2).

2. A high degree of consistency in response to group pressure was found within a given task at a single sitting; Asch's finding that *Ss* who were independent early in the experimental session continued independent and *Ss* who yielded early in the experimental session continued to yield is substantiated with the present population.

3. Response to group pressure was significantly consistent between tasks at a single sitting. The *Ss* who were low yielders on one task tended to be low yielders on other tasks; high yielders on one task tended to be high yielders on other tasks. Phi coefficients ranging from .33 to .54 were obtained between pairs of tasks with respect to high and low yielders.

4. Response to group pressure was significantly consistent over a period of time. The *Ss* who yielded or did not yield during one sitting tended to yield or not to yield, respectively, during later sittings as shown in Table 1.

5. The embedded figures test (4, 5) was administered to 20 high and 20 low yielders to explore the relationship between this test and the group pressure situation. No significant difference was found between the two groups in the amount of time they required to locate the simple in the

TABLE 1

CONSISTENCY OF YIELDING BEHAVIOR OVER A TIME LAPSE

Subjects	No. of Judgments Showing Yielding	
	Lines Session I*	Lines Session II
Yielders		
1	12	11
2	8	5
3	8	12
4	11	1
5	9	12
Total	48	41
Nonyielders		
1	0	1
2	0	0
3	2	1
4	2	0
5	0	1
Total	4	3

* Procedure identical to that described by Asch (2).

[1] This study is part of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy to the Graduate Faculty of Political and Social Science, The New School for Social Research, New York City. The writer owes a special debt of gratitude to Dr. Mary Henle for her encouragement of the research and for her critical reading of the manuscript.

[2] Thanks are owed to Miss Evelyn Lehrer, Assistant Director (formerly) of the Westchester School of Nursing and to Miss Barbara Tate, Associate Director of St. Lukes Hospital School of Nursing for helping to make their respective student bodies available.

[3] Asch (2) has shown that the effect of a majority of three on the judgments of naive *Ss* is as great as that of larger groups.

[4] For details, the reader is referred to the dissertation (3).

complex figures, though the existing differences were in the expected direction. The yielders failed to locate the simple in the complex figures significantly more often than the nonyielders.

REFERENCES

1. ASCH, S. E. *Social psychology*. New York: Prentice Hall, 1952.
2. ASCH, S. E. Studies on independence and conformity [PT. 1], a minority of one against a unanimous majority. *Psychol. Monogr.*, 1956, **70**, No. 9 (Whole no. 416).
3. ROSNER, S. Studies of group pressure. Unpublished doctor's dissertation, New School for Social Research, 1956.
4. WITKIN, H. A. Individual differences in ease of perception of embedded figures. *J. Pers.*, 1950, **19**, 1-15.
5. WITKIN, H. A. *et al. Personality through perception*. New York: Harper, 1954.

Received September 28, 1956.

ON THE RELATIONS BETWEEN DISCRIMINATION REACTION TIME, ANXIETY, AND INTELLIGENCE

G. ROBERT GRICE

University of Illinois

IN A test of a theory concerning the effects of drive on performance involving conflicting habits, Grice (4) found the predicted difference on a complex discrimination reaction time task between two groups differing in "manifest anxiety." However, a covariance analysis indicated that this difference could be attributed to differences between the groups in intellectual ability. Freeman and Maher (3) have criticized Grice's conclusion on the basis of speculations that the Discrimination Reaction Time Test (DRT) is "largely perceptual in character" and that perceptual factors in the Clerical Aptitude Index (employed as an intellectual measure) may have produced the obtained relation rather than general intellectual ability. Citing the excellent study by Fleishman and Hempel (2) as evidence, they argue that intellectual components in DRT tend to disappear with practice and that perceptual speed and simple reaction time assume greater importance.

In the first place, it should be pointed out that the task analyzed and described by Grice was quite different from that employed by Fleishman and Hempel. Instead of a long period of practice with a single set of instructions, the task (Grice's Problem 5) involved a signal, one second in advance of each reaction, which informed *S* which of four rather complicated sets of instructions were to be employed in determining the correct response. It would appear likely that this involved more intellectual ability than even the early trials of the conventional DRT task. In the second place, it should be pointed out that "Perceptual Speed" was the only purely perceptual factor which emerged in the Fleishman and Hempel analysis and that it accounted for only five per cent of the DRT variance even after long practice. This cannot be regarded as strong evidence that the test was "largely perceptual in character."

The other point raised by Freeman and Maher concerned the suitability of the Clerical Aptitude Index as a general intellectual measure and the question of the presence of perceptual factors in it. They ask, rhetorically, "Were not Air Force data on tests more universally recognized as general intellectual measures available?" The answer is simple. No. The Clerical Aptitude Index is actually a rather good measure of intellectual ability as Grice originally stated. It is based on four tests: Word Knowledge, a test of academic vocabulary; Background for Current Affairs, a test of general information; Numerical Operations, a test involving simple computations; and Dial and Table Reading, a test requiring the verification of readings on groups of dials and the determination of certain information by reading mathematical tables. It was this last test which lead to the statement by Grice that the index contained perceptual factors. However, perceptual ability is by no means its major source of variance. In a factor analysis by Zachert (5), the test had a loading of .33 on a factor named Perceptual Speed. It also had loadings of .50 on Numerical Facility, .39 on General Reasoning, and .29 on Verbal Comprehension. In a more recent analysis by Bechtoldt (1), it had a loading of .17 on a Space factor and loadings of .31 and .26 on Academic Background and Induction respectively. The other tests show their major loadings on factors named Verbal Comprehension, Number, and Academic Background (1). All four tests have their major projections on second-order factors labeled by Bechtoldt as "G" (General Intellective) and "VM" (Facility with Verbal Materials).

With the nature of the tests in mind, it can now

TABLE 1
PERFORMANCE OF HIGH AND LOW ANXIOUS GROUPS ON EACH TEST

Test	High A Mean	LowA Mean	t	p
Word Knowledge	4.40	6.03	3.47	$<.01$
Numerical Operations	4.40	5.37	2.05	$<.05$
Dial & Table Reading	3.37	4.70	2.51	$<.02$
Bkgd. Cur. Aff.	4.13	5.40	2.61	$<.01$

be pointed out that in Grice's study, each of the four significantly differentiated between the high and low anxious groups and that each was also positively correlated with response speed in the reaction time task. The means of the high and low anxious groups and the values of t are presented in Table 1. The product-moment correlations of response speed in Problem 5 in the DRT task with the individual tests were as follows: with Word Knowledge, .50; with Numerical Operations, .42; with Dial and Table Reading, .63; and with Background for Current Affairs, .33.

These considerations strongly suggest that the relationships reported by Grice were due to intellectual variance in common with "manifest anxiety," "clerical aptitude," and the reaction time task, and that the speculations of Freeman and Maher are contrary to fact.

REFERENCES

1. BECHTOLDT, H. P. Factor analysis of the Airman Classification Battery with civilian reference tests. USAF, Hum. Resour. Res. Cent., *Res. Bull.*, 1953, No. 53-59.
2. FLEISHMAN, E. A., & HEMPEL, W. E. The relation between abilities and improvement with practice in a visual discrimination reaction task. *J. exp. Psychol.*, 1955, **49**, 301-312.
3. FREEMAN, J. T., & MAHER, H. A note on "Discrimination reaction time as a function of anxiety and intelligence." *J. abnorm. soc. Psychol.*, 1957, **54**, 134-135.
4. GRICE, G. R. Discrimination reaction time as a function of anxiety and intelligence. *J. abnorm. soc. Psychol.*, 1955, **50**, 71-74.
5. ZACHERT, VIRGINIA. Factor analyses of the Army, Navy, and Air Force classification batteries. USAF, Hum. Resour. Res. Cent., *Res. Bull.*, 1952, No. 52-12.

Received March 7, 1957. Prior publication.

SECURITY OF JUDGES AS A FACTOR IN IMPRESSIONS OF WARMTH IN OTHERS

JOSEPH BOSSOM AND A. H. MASLOW
Brandeis University

THE effect of variation among the people being judged has been considered in previous studies of first impressions (2). The present study is concerned with the effect of variation among judges upon their first impressions of others. Differences in the first impressions of secure and insecure judges are investigated.

METHOD

Subjects

Judges were 105 undergraduate students enrolled in an elementary course in psychology. From this total group were selected 22 judges who made scores of 9 points or less on the Maslow S-1 test for security (1), the secure group. Their scores ranged from 0.89 to 1.62 *SD* below the mean of the total group. Twenty-two judges, who scored 30 points or more comprised the insecure group. Their scores ranged from 0.71 to 2.65 *SD* above the total group mean.

Stimulus

The "objects" to be judged consisted of the first 200 photographs obtained from yearbooks of former graduating classes. Except for difference in age (about two years), the "objects" and the judges come from about the same socioeconomic, racial, and religious background.

Exposure to Stimulus

Each judge was handed an answer sheet with line space for 200 responses. The categories of response were headed: Very Warm; Warm; Cold; Very Cold. All judges were given the following verbal instructions:

> You will be shown 200 photographs. They will appear on the screen in front of the room. These photographs will be changed rather rapidly. After looking at a photograph you will have sufficient time to indicate your impression of the person in the photograph. This is done by placing a mark in the appropriate box on your answer sheet.

Subsequently, the 200 photographs were projected on a screen. Each photograph was exposed for about 5 seconds. Periodically, the number of the photograph was called out so that judges could determine whether they had missed an item.

RESULTS AND DISCUSSION

Judges in all groups were more likely to ascribe warmth than coldness to the persons represented in the photographs. The secure group, the insecure group, and the remaining judges (others) were compared with respect to the proportion in each group who ascribed warmth more frequently than coldness. As can be seen in Table 1, the secure

TABLE 1
PERCENTAGE OF JUDGES IN EACH GROUP ASCRIBING WARMTH MORE OFTEN THAN COLDNESS TO PHOTOGRAPHS

Response Categories Compared	Group			Probability of Difference	
	Secure (N = 22)	Insecure (N = 22)	Others (N = 61)		
Warm *vs.* Cold	95%	68%	78%	Sec. *vs.* Others	.10
				Insec. *vs.* Others	.50
				Sec. *vs.* Insec.	.05
V. Warm & Warm *vs.* V. Cold & Cold	90%	54%	78%	Sec. *vs.* Others	.80
				Insec. *vs.* Others	.05
				Sec. *vs.* Insec.	.02
V. Cold *vs.* V. Warm	77%	68%	42%	Sec. *vs.* Others	.01
				Insec. *vs.* Others	.05
				Sec. *vs.* Insec.	.70

and insecure groups differed significantly (p = .05) in the number of judges scoring Warm as compared with Cold. The difference between these groups was also significant when the Very Warm and Warm categories taken together were compared with Cold and Very Cold (p = .02). In these comparisons a greater percentage of judges gave Warm responses in the secure group than among the "others." The "others," in turn, had a greater percentage of judges scoring warmth than the insecure group. When only the Very Warm and Very Cold categories were considered, significant differences were not obtained between the secure and insecure groups. In addition, a higher percentage of judges in both extreme groups employ the Very Warm category than among the intermediate group of "others."

Since our data do not show whether the faces in the photographs are in "reality" warm or cold, we cannot attribute the obtained group differences to differential efficiency of perception of reality, projections based upon habitual expectations in interpersonal relationships, or to Freudian projection. Decision among these alternatives must await further research.

SUMMARY

From a larger group of judges, a secure and an insecure group were selected. All judges were exposed to a series of 200 photographs. Judges were required to indicate whether the person in the photograph was Very Warm, Warm, Cold, or Very Cold. The percentage of judges reporting Warm impressions more often than Cold impressions tended to be greater in the secure group than in the insecure group.

REFERENCES

1. MASLOW, A. H. A clinically derived test for measuring security-insecurity. *J. gen. Psychol.*, 1943, **33**, 21-41.
2. TAFT, R. The ability to judge people. *Psychol. Bull.*, 1955, **52**, 1-21.

Received August 27, 1956.

The Basis of Selective Attraction

The Psychology of Sexual Emotion

By VERNON W. GRANT

This remarkable book brings together for the first time in organized form the best that has been written on the emotional and aesthetic side—as distinguished from the biological side—of sexual attraction and attachment.

It is uniquely adapted to the needs of college courses in marriage psychology and mate selection, and will be of value to marriage counselors and to clinical psychologists in dealing with certain problems of sex maladjustment.

Dr. Grant, who is head of the Department of Psychology, Hawthornden State Hospital, Macedonia, Ohio, presents and critically discusses the varying definitions of sexual love, the state known as "being in love," the phenomenon of choice, or selective attraction, the relation of sexual aesthetics to general aesthetics, and the influence of cultural factors, such as the status of women.

Among the writers whose opinions and observations are quoted are Proust, Santayana, Zola, Flaubert, H. G. Wells, and Stendhal; among the psychologists are Albert Moll, Havelock Ellis, Freud, Jung, Magnus Hirschfeld, Stekel, Krafft-Ebing, and Theodore Reik. Some of the material has never before been available in English translation.

At all bookstores • $4.75

LONGMANS, GREEN AND COMPANY

55 Fifth Avenue, New York 3